The interplay of global chromosomal organisation, promoter-enhancer interactions and transcription

Michiel Johannes Thiecke

The Babraham Institute

Saint Edmund's College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2019

I just wondered how things were put together

– Claude E. Shannon (1916 - 2001)

Declaration

This dissertation is the result of my own work and it includes nothing that is the outcome of work done in collaboration, except where specifically indicated.

This thesis is not substantially the same as any that I have previously submitted, or, is currently being submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution. I further declare that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other university or similar institution except as specified in the text.

This thesis does not exceed 60,000 words, excluding figures, the bibliography and the appendix.

Michiel Johannes Thiecke

September 2019

Summary

All somatic cells within an organism contain the same genetic material, yet they display pronounced differences in function and morphology. Precise control of gene expression is of fundamental importance to allow cells to properly develop, maintain homeostasis, and respond to external stimuli. The first step in gene expression is transcription, which starts at the core promoter region. While core promoters are crucial for transcriptional initiation, they are insufficient for establishing complex tissue- and condition-specific gene expression patterns in multicellular organisms. Additional transcriptional control elements, such as gene enhancers, are required for this, with many such elements localising considerable distances away from their target promoters. Enhancers commonly convey their regulatory signals to target promoters by forming physical contacts with them through three-dimensional DNA looping, underpinning the importance of chromosomal organisation in transcriptional control. In recent years, the emergence of chromosome conformation capture and related methodologies has dramatically increased our understanding of chromosomal organisation. In particular, high-throughput Hi-C analyses across cell types have led to the identification of spatial genomic structures, including Topologically Associating Domains (TADs). In parallel, high-resolution versions of these technologies (such as 5C, CHiA-PET, HiChIP and Capture Hi-C) have detected multitudes of novel looping interactions, including connections between promoters and enhancers. The interplay between precise regulatory interactions, the higher-order chromosomal organisation, and their joint contribution to transcriptional control is incompletely understood and is the focus of this work.

In the first part of this work, I take advantage of high-resolution Promoter Capture Hi-C (PCHi-C) data to investigate the localisation of promoter interactions with respect to TAD boundaries in human primary blood cells and cell-cycle synchronised HeLa cells. I show that the majority of promoter interactions originate at, and are constrained by TAD boundaries. However, a minority of promoter interactions appear to cross TAD boundaries in all analysed cell types. Furthermore, I identify genes with multiple TAD-boundary crossing interactions per promoter and present evidence that these interactions may be supported by transcriptional machinery. These results

suggest a role for transcriptional machinery in shaping promoter interactions in a TAD independent manner.

In the second part of this work, I investigate promoter interaction rewiring upon perturbations of architectural proteins. For this analysis, I use PCHi-C data from HeLa cells, in which cohesin or CTCF are rapidly depleted using Auxin-induced degradation. I show that promoter interactions that are lost, maintained, or gained upon cohesin depletion possess distinct distance profiles and relate to TAD organisation in markedly different ways. I demonstrate that promoter-interacting regions that are lost upon cohesin depletion associate with architectural proteins, while those that are maintained or gained show characteristics of enhancers. Finally, I show evidence for a functional role of cohesin-mediated interactions in transcriptional regulation.

Collectively, this work reveals the interplay between TADs, promoter interactions and transcription, while suggesting that promoter interactions may be supported by TAD independent mechanisms.

Acknowledgements

Over the past four years I have been supported professionally as well as personally by many people. Without them, this project would have been a lot more difficult and a lot less enjoyable.

Firstly, I want to thank my supervisor Mikhail Spivakov for giving me the opportunity to do this PhD. Additionally, I want to mention that your help and feedback were invaluable and that your enthusiasm is nothing short of infectious. Additionally, I want to thank Anne Corcoran, who became my official supervisor for the last year of my PhD.

I also want to thank the other members of the Spivakov group, in particular Lina Dobnikar. It was a great experience to work together, not in the least when we managed to win the Biotech YES award. I also want to mention Jonathan Cairns who was my mentor for the first two years of my PhD and with whom I had many fun and stimulating conversations about math, science and comedy. Then, I want to thank Jörg Morf for the many scientific discussions and social conversations we had.

I want to thank my friends from Milton road, you are all very dear to me and I look forward to reuniting in the future to watch bad movies, cook, eat and have fun together.

Then, I want to mention my friends back in the Netherlands (de vriendjes). Your friendships have been a great support throughout this project and I cannot wait to see you all again.

I want to thank my parents who have been very supportive and loving. Also, your help was indispensable, especially towards the end of this project.

I also want to thank my sister Imktje and her partner Jeroen, you are very dear to me and it is always a lot of fun to meet with you and your two boys.

Last but certainly not least, I want to thank my girlfriend Gwen Moonen, who has lovingly supported me throughout my PhD, first from the Netherlands and later when she joined me in Cambridge. You always make me smile and I look forward to our future together.

Acknowledgement of assistance

1) Initial training techniques and subsequent mentoring:

Dr Jonathan Cairns – Mentor

Dr Valeriya Malisheva – Mentor

Dr Peter Rugg-Gunn – Assessor

Dr Chris Wallace – University supervisor

Dr Mikhail Spivakov – Advice on data analysis and interpretation

Dr Steven Wingett – Advice on data analysis and software use

Dr Simon Andrews – Advice on data analysis and software use

Dr Anne Segonds-Pichon – Advice on statistical analyses

2) Data obtained from a technical service provider

The Babraham sequencing facility – Bioanalyzer quality control and Illumina sequencing

The IMP sequencing facility – Bioanalyzer quality control and Illumina sequencing

3) Data produced jointly

Dr Valeriya Malisheva – ATAC-seq experiments

4) Data/materials provided by someone else

Biola Maria Javierre – Hi-C and PCHi-C in haematopoietic tissues

Dr Csilla Varnai – TAD calls on haematopoietic lineages and in HeLa

Dr Gordana Wutz – ChIP-Seq, RNA-Seq, Hi-C, and PCHiC in HeLa

Dr Stefan Schoenfelder – Hi-C, PCHiC in HeLa

Dr Wen Tang – qPCR on CRISPR-dCAS9-KRAB targeted genes

Table of contents

Introduction

Gene regulation lies at the basis of cell diversity, defining the wide range of morphologically distinct states that cells display. Yet, genetic information is the same in all somatic cells within an organism. Since genetic misregulation is involved in many diseases and syndromes, identifying mechanisms that govern gene regulation advances fundamental as well as clinical research fields such as developmental, synthetic, and cancer biology. Transcriptional regulation forms the first step in the sequence of mechanisms that precisely tune the abundance of all gene products. Therefore, it is arguably one of the most crucial gene regulatory mechanisms. Transcriptional regulation is dependent on a wide range of cellular and nuclear processes, among which spatial genomic organisation has emerged as a major influence in eukaryotic cells. Furthering our understanding of spatial genome organisation holds great promise for understanding a wide range of pathologies. For example, (Valton et al. 2016) showed that somatic mutations may drive alterations to spatial genome organisation leading to oncogenesis. Such findings provide novel insights into the mechanisms that cause disease and may lead to improved prevention and treatment plans.

1.1 Spatial organisation of the eukaryotic genome

The eukaryotic genome is composed of chromosomes which are tightly packaged within the nucleus. For example, the human genome typically has 23 pairs of chromosomes which, if unfolded and placed contiguously, would amount to roughly 2 metres in length. Since the diameter of the nucleus is on the micrometre scale, DNA must be folded to fit inside of it. Following the mitotic telophase, DNA de-condenses and expands throughout the nucleus. Although interphase chromosomes had been considered to be randomly distributed throughout the nucleus, in recent decades it has become clear that this is not the case. There are different levels of chromosomal organisation, including chromosome territories (CTs), euchromatin/heterochromatin, A/B compartments, topologically associating domains (TADs) and precise DNA interactions between loci such as promoters and enhancers.

1.1.1 Chromosome territories

Non-mitotic chromosomes occupy chromosome territories (CTs): distinct regions within the nucleus, where distant loci on the same chromosome tend to be spatially more proximal than loci on different chromosomes (Cremer *et al.*, 2001, 2010). CTs were discovered through laser-UV micro-irradiation (Cremer *et al.*, 1982), subsequently detected through fluorescent in situ hybridization microscopy (Schardin *et al.*, 1985) and recently confirmed by sequencing-based polymer modelling (Dekker *et al.*, 2013). Although CTs might suggest a static nuclear organisation where genes are restricted to their CTs, interactions between territories have been shown to occur (Branco *et al.*, 2006). For example, in the context of development, genes have been shown to interact with DNA elements in other chromosome territories, by relocating to the CT boundary and even by looping out of their CT (Chambeyron *et al.*, 2004; Chaumeil *et al.*, 2006). Furthermore, CT positioning and pairing may be tissue-specific (Cremer *et al.*, 2001). Lastly, reshuffling of lamina-associated domains upon cell division (Kind *et al.*, 2013) further supports the notion that CTs may change position and pair up with different neighbouring CTs.

Collectively, CTs are a high-level feature of nuclear architecture which shows that DNA is spatially organised. This organisation restricts the chromosome polymers within the nucleus, while allowing for dynamic rearrangement, such as in development.

1.1.2 Euchromatin and heterochromatin

The combination of DNA and DNA-bound proteins is called chromatin and the association with proteins plays a major role in shaping DNA conformation. Microscopic analyses have shown that DNA separates into two broad categories: euchromatin and heterochromatin. Heterochromatin is transcriptionally less active than euchromatin, it is typically located at the nuclear periphery and it may associate with the nuclear lamina (Huisinga *et al.*, 2006; Solovei *et al.*, 2013). However, genes residing in euchromatic regions may also be transcriptionally silent and genes residing in heterochromatic regions may be transcriptionally active (Huisinga *et al.*, 2006; Azzalin *et al.*, 2007). A major difference between the two states lies in their compaction which arises from differential nucleosome density. Nucleosomes consist of two of each of the histone proteins H2A, H2B, H3, and H4, which form an octamer

around which DNA is wrapped in lengths of 146 base pairs (bp). Nucleosome-dense chromatin is tightly compacted and transcriptionally less active than nucleosome-sparse chromatin, which is more readily transcribed. Furthermore, covalent modifications of the terminal regions of histone amino acid polymers are related to nucleosome density as well as differential transcriptional activity (Bannister *et al.*, 2011). Many histone modifications have been identified, with varying and often strongly contextual roles. For example, histone H3 trimethylation of lysine residues 27 (H3K27me3) and H3K4me3 are related to gene inactivity and activity respectively, and are thought to play a role in conditionally silencing or activating genes in the context of differentiation and development (Vastenhouw *et al.*, 2012).

In addition, genes located in euchromatic regions are frequently marked by H3K4me3 and H3K36me3 whereas distal regulatory elements commonly bear H3K4me1 (Bannister *et al.*, 2011) (For a discussion of distal elements see section 1.2.4). Heterochromatin is further subdivided into two groups: constitutive and facultative. Constitutive heterochromatin, such as that found in the telomeric and centromeric regions of the chromosomes, is most strongly associated with the nuclear peripheral space and is most consistently transcriptionally inactive, containing only a handful of genes (Saksouk *et al.*, 2015). In contrast, facultative heterochromatin is farther removed from the periphery, is more gene-rich and the enclosed genes may be transcriptionally activated. For instance, genes involved in differentiation and developmental processes may temporarily activate, followed by silencing upon progression of the governing processes. Heterochromatin is epigenetically marked by global histone hypoacetylation (Saksouk *et al.*, 2015) and local DNA hypermethylation (Bird, 2002). Constitutive heterochromatin is canonically marked by H3K9me3 and high levels of Heterochromatin Protein 1, while facultative heterochromatin is typically marked by H3K27me3 and Polycomb repressor complexes (Bannister *et al.*, 2011). Interestingly, the association between heterochromatin and the nuclear periphery is inverted in photoreceptor rod cells of nocturnal mammals (Solovei *et al.*, 2009), and it spontaneously inverts upon disruption of lamina association (Solovei *et al.*, 2013).

Collectively, the complex and dynamic spatial organisation of chromatin shows that the three-dimensional organisation of DNA is inherently related to gene activity.

1.1.3 A/B compartments

In recent years, sequencing-based strategies for studying chromosomal organisation such as Hi-C (see section 1.3.2) have dramatically increased our understanding of spatial genome organisation. Notably, eigenvector decomposition of pairwise DNA interactions revealed that the genome separates into two compartments (A and B), with loci interacting with higher frequency within, as compared to between, compartments (Lieberman-Aiden *et al.*, 2009). Upon discovery, chromatin residing in the A-compartment was shown to be more transcriptionally active and more susceptible to DNAse1 degradation, suggesting that the A-compartment is less compacted. Initially, Lieberman-Aiden et al. showed compartmentalization in the highly rearranged genome of K562 cells. Single cell Hi-C analysis in mouse T-cells has confirmed A/B compartmentalization and supported the notion that A-compartments show characteristics of euchromatin, while B-compartments show characteristics of heterochromatin (Nagano *et al.*, 2013).

Taken together, it is likely that A/B compartmentalization represents a complementary view on the separation between euchromatin and heterochromatin. This indicates that sequencing-based analysis techniques are capable of detecting previously identified structures, in addition to providing a higher resolution and throughput than attainable through conventional microscopy.

1.1.4 Topologically associating domains

It has long been proposed that eukaryotic DNA is organised into higher-order domains that are involved in crucial processes such as chromatin compaction and insulation of regulatory influences (Kellum *et al.*, 1991, 1992; Dillon *et al.*, 2000). However, until recently, such structures had eluded detection. Enabled by increased sequencing capabilities and reduced cost, higher resolution chromatin conformation capture contact maps enabled detection of increasingly finer DNA structures. This led to the discovery of topologically associating domains (TADs): regions of the genome that insulate DNA-DNA interactions across their boundaries and that show preferential self-interaction within their boundaries (Dixon *et al.*, 2012; Nora *et al.*, 2012) (Figure 1). On the linear genome, TADs are between hundreds of kbp to several Mbp in length (Dixon *et al.*, 2012) and are thought to exist as a hierarchy of meta-TADs and sub-TADs (Fraser *et al.*, 2015; Zhan *et al.*, 2017). TADs are

frequently conceptualised as a series of nested loops that are demarcated by the architectural proteins cohesin and CCCTC-binding factor (CTCF) (Figure 1). The latter binds to specific DNA sequences called CTCF binding sites, which are frequently found at TAD boundaries. TADs show considerable conservation between tissues as well as species. Exploiting this property, (Rudan et al. 2015) have shown that genetic divergence of individual CTCF binding sites at TAD boundaries is insufficient for TAD boundary disruption, indicating that TAD boundaries are likely supported by multiple redundant elements which strengthen conservation of TAD organisation. The fact that TADs show considerable conservation between tissues as well as between species suggests that they play a non-trivial role in spatial genome organisation (Jin *et al.*, 2013; Rao *et al.*, 2014; Vietri Rudan *et al.*, 2015). TADs have been implicated in various biological functions such as DNA replication (Uusküla-Reimand et al. 2016), replication timing (Pope et al. 2014), chromatin unknotting (Racko et al. 2018), and transcription (see sections 1.1.4.4 and 1.1.4.5). However, their functional roles and even existence are subject to ongoing study and debate. Since TADs are detected by analysis of the genomic organisation of many cells at a time, the possibility exists that the apparent TAD structures on DNA contact maps represent an emergent property of conflated data. If that is the case, TADs might not exist as stable entities within single cells, but rather, represent genomic regions where DNA interactions have a higher probability of forming or persisting. Single-cell chromatin conformation capture approaches are limited by DNA contact sparsity, preventing reliable detection of TADs. However, single-cell imaging approaches appear to be approaching the capacity to resolve TADs, reporting domain-like nanocomparments in repressed chromatin in *Drosophila melanogaster* (Boettiger *et al.*, 2016; Szabo *et al.*, 2018) and in non-repressed chromatin in human cells (Bintu *et al.*, 2018). The latter reported TAD-like structures with variable boundary positions, but with a preference for locations where cohesin and CTCF are present. Collectively, these findings show that on the scale of hundreds of kilobase pairs (kbp) to millions of base pairs (Mbp), DNA interactions appear to be organised in TADs. TAD boundaries associate with cohesin and CTCF and show a capacity to confine DNA interactions within their boundaries. Therefore, emerging evidence suggests that TADs may exist stably within single cells.

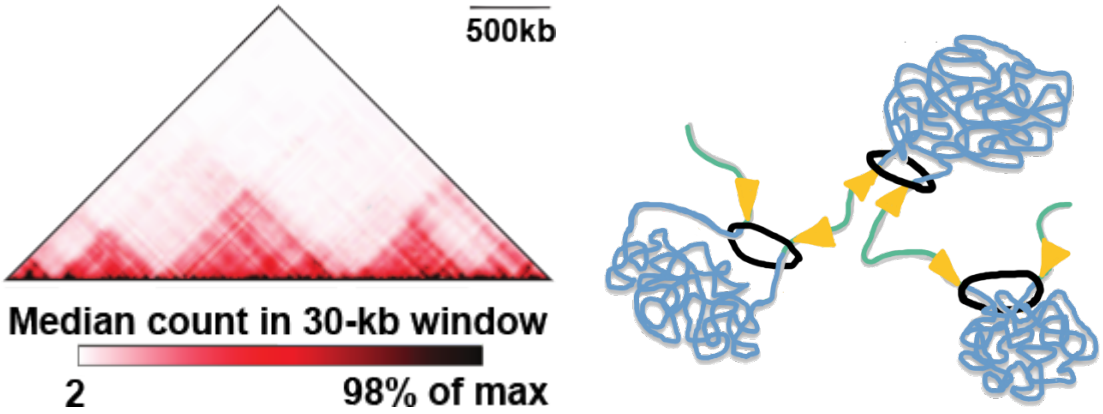

Figure 1. Topologically associating domain detection and conceptualization. Left: a 5C heat map showing TADs on the X-chromosome of undifferentiated mouse embryonic stem cells at a 30kb resolution. TADs are regions with higher interaction frequencies within their boundaries as opposed to across them and they appear as triangles in the heat map. TADs span scales on the order of 10^5 to 10^6bp and they exhibit apparent hierarchical organisation. Note that the TAD heat map represents a population signal on the order of 10^7 cells. Right: a conceptual illustration of potential TAD organisation at the level of an individual DNA strand, including the architectural proteins cohesin and CTCF (blue: TADs, green: non-TAD regions, black rings: cohesin, yellow triangles: CTCF). Note that CTCF is shown in convergent orientation. Figures adapted from: (Nora *et al.*, 2012; El-Sharnouby *et al.*, 2017).

1.1.4.1 TADs and A/B compartmentalisation

TADs and A/B compartmentalisation likely represent largely independent aspects of genome organisation. This is supported by a number of observations: mouse maternal zygotes have been shown to lack compartmentalization while TADs are present (Flyamer *et al.*, 2017). In contrast, upon cohesin depletion, compartmentalization is largely preserved (Seitan *et al.*, 2013) or even mildly strengthened, whereas TAD organisation strongly reduces (Haarhuis *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017; Wutz *et al.*, 2017; Nuebler *et al.*, 2018), see Figure 2. Compartment strengthening in the absence of TADs led to the suggestion that TAD organisation exhibits a capacity to counteract A/B compartmentalisation, possibly by preventing the spread of heterochromatin across TAD boundaries (Nora *et al.*, 2012; Dixon *et al.*, 2015), although CTCF binding has been suggested to lack

this property (Nora *et al.*, 2017). Interestingly, localized induction of repressive chromatin modifications H3K27me3 and H3K9me3 showed DNA interaction rewiring consistent with compartment switching at the sub-TAD level (Wijchers *et al.*, 2016), suggesting sub-TAD organisation may halt the spread of heterochromatin. However, a separate analysis showed that Polycomb repressive complex (PRC)-associated TADs can reside within the A compartment (Moore *et al.*, 2015), suggesting that transcriptional silencing can occur on the TAD level, without the need for compartment switching. Lastly, upon nuclear inversion by disruption of lamina association, TAD organisation appears largely unaffected (Falk *et al.*, 2018), showing that TADs persist even upon large scale relocation of peripheral and central chromatin. Together, these findings support the notion that TAD organisation is largely independent from compartmentalisation, although TAD organisation may counteract compartmentalization to some extent.

1.1.4.2 TADs and cohesin

The cohesin complex was initially known for its involvement in sister chromatid cohesion and separation upon transitioning from the mitotic metaphase to the anaphase (Nasmyth *et al.*, 2009). However, cohesin has also been implicated in additional aspects of spatial genome organisation throughout the cell cycle, including regulation of transcriptional activity by supporting promoter-enhancer interactions (Rollins *et al.*, 2004; Dorsett *et al.*, 2005) and by extension, development (Hadjur *et al.*, 2009). Cohesin consists of four core subunits: two SMC subunits, a kleisin subunit and a HEAT repeat subunit. Cohesin proteins are strongly conserved in eukaryotes. In humans, these proteins are called Smc1/Smc1β, Smc3, Scc1/Rec8, and SA1/SA2/SA3. (Peters *et al.*, 2008; Nasmyth *et al.*, 2009). The first three proteins form a ring-like structure which binds DNA in a sequence independent fashion and may relocate along the DNA strand. It is commonly accepted that cohesin rings surround DNA strands, since disruption of the cohesin ring by protein degradation has been shown to completely remove cohesin binding to DNA in vivo (Gruber *et al.*, 2003). In the context of loop extrusion, the question whether cohesin relocates by active, ATP-driven, processes or if it is passively being moved along the DNA by other genomic proteins is not fully established. The cohesin-related protein complex condensin has recently been shown to be capable of ATP-dependent loop extrusion in vitro (Ganji *et al.*, 2018), suggesting that cohesin might operate similarly.

However, studies in yeast have shown that cohesin is highly enriched at intergenic sites of convergent transcription (Glynn *et al.*, 2004; Lengronne *et al.*, 2004; Betts Lindroos *et al.*, 2006), suggesting that RNA polymerases may be driving cohesin relocation. Although this enrichment is not observed in human cells, perhaps owing to the large size of intergenic regions (Parelho *et al.*, 2008; Wendt *et al.*, 2008), recent work suggests that cohesin is relocated by transcription (Busslinger *et al.*, 2017). Taken together, although the mechanism that drives cohesin positioning is disputed, it is clear that cohesin binds to DNA, tethers DNA and relocates along DNA strands.

1.1.4.3 Loop extrusion

Polymer modelling studies have suggested that TAD formation may be driven by a process called loop extrusion (Alipour *et al.*, 2012; Sanborn *et al.*, 2015; Fudenberg *et al.*, 2016; Nuebler *et al.*, 2018). Loop extrusion describes a process where a loop of DNA is formed by an extruding factor that relocates along the DNA, thereby enlarging the loop. Cohesin has been suggested as the pivotal loop extruding factor in this process (Alipour *et al.*, 2012). The loop extrusion model stipulates that the extrusion process halts upon encountering TAD boundaries as a result of interacting with TAD boundary associated proteins (Sanborn *et al.*, 2015; Wit *et al.*, 2015; Fudenberg *et al.*, 2016). The zinc finger protein CTCF is thought to perform this function, specifically when bound to DNA in a convergent orientation (Figure 1) (Rao *et al.*, 2014; Sanborn *et al.*, 2015). CTCF is known as a transcriptional insulator (Filippova *et al.*, 1996; Bell *et al.*, 1999) which binds to highly specific DNA motifs which are enriched at TAD boundaries (Dixon *et al.*, 2012; Rao *et al.*, 2014). Several studies have confirmed that the effects of deleting or inverting CTCF binding sites rearranges chromatin loops (Guo *et al.*, 2015; Narendra *et al.*, 2015; Sanborn *et al.*, 2015; Wit *et al.*, 2015) and that near-complete CTCF depletion shows a dramatic reduction in TAD organisation (Nora *et al.*, 2017), see Figure 2. However, the majority of CTCF binding sites are not located at TAD boundary regions (Dixon *et al.*, 2012). Additionally, many TADs show no enrichment for CTCF at their boundaries, instead showing typical marks of transcriptional activity (Bonev *et al.*, 2017). Consistent with the loop extrusion model, cohesin depletion shows major loss of TAD organisation in bulk population analyses (Sanborn *et al.*, 2015; Fudenberg *et al.*, 2016; Wutz *et al.*, 2017), while single cell analyses show that upon cohesin

depletion, DNA interactions are less likely to be positioned at cohesin/CTCF sites (Bintu *et al.*, 2018). This confirms that cohesin plays a major role in shaping spatial DNA organisation. Furthermore, it supports the notion that cohesin plays a pivotal role in shaping domain structures by tethering (distant) genomic loci. Although DNA interactions do appear to exist in the absence of cohesin, their confinement within TADs appears supported by cohesin. In further support of the loop extrusion model, increase of cohesin retention time by disruption of the cohesin dissociation factor Wapl results in cohesin accumulation on the DNA and subsequently, a semi condensed genome called the “vermicelli” phenotype (Tedeschi *et al.*, 2013; Haarhuis *et al.*, 2017; Wutz *et al.*, 2017). This suggests that chromatin condensation is governed by parameters such as association and dissociation rates of architectural proteins, which is supported by polymer simulations of DNA condensation through loop extrusion (Goloborodko *et al.*, 2016). In a recent polymer modelling analysis, it has been suggested that loop extrusion may be responsible for active unknotting of interphase DNA (Racko *et al.*, 2018). Collectively, these findings suggest that TADs may be formed through loop extrusion, that cohesin likely supports this process and the resulting spatial organisation, and finally, that loop extrusion halts at CTCF binding sites, possibly among other barriers. The loop extrusion model offers an elegant explanation for TAD formation as well as chromatin unknotting although it does not exclude the possibility of additional processes that shape DNA interactions.

1.1.4.4 Properties of TAD boundaries

Properties of TAD boundaries have been investigated in order to attribute potential functions to TADs. TAD boundaries have been shown to be enriched for the canonical promoter mark H3K4me3 in addition to transcription start sites (TSSs) and nascent transcripts, while showing depletion for the constitutive heterochromatin associated mark H3K9me3 (Dixon *et al.*, 2012). This suggests that TAD boundaries are sites of active transcription. Interestingly, the enhancer associated mark H3K4me1 does not appear to be enriched at TAD boundaries (Dixon *et al.*, 2012), suggesting that TAD boundaries do not play a direct role in promoter-enhancer pairing. TAD boundaries have also been shown to associate with type II topoisomerase (Uusküla-Reimand *et al.*, 2016), which is involved in relieving mechanical stress on DNA as a result of supercoiling. Such stress is formed by

processes such as transcription, chromatin compaction/decompaction and replication. Involvement of TADs in the latter is supported by the observation that replication domain boundaries and TAD boundaries are often found in the same place on the genome (Pope *et al.*, 2014). Lastly, TAD boundaries have been shown to be enriched for the histone modification H3K36me3 (Dixon *et al.*, 2012), which is suggested to be related to mRNA splicing (Schwartz *et al.*, 2009). Collectively, TADs appear to be involved in various processes while offering an exciting mechanism for supporting transcriptional regulation by distal regulatory elements.

1.1.4.5 TADs and promoter-enhancer interactions

Upon their discovery, TADs were suggested as a mechanism for forming and facilitating promoter-enhancer interactions (see section 1.1.5) (Nora *et al.*, 2012). In this model, the loop-like nature of TADs provides an environment within which promoters and enhancers (see sections 1.2.1 and 1.2.4.2) are brought within proximity of each other whilst the insulating properties of TAD boundaries restrict promoter enhancer interactions from forming ectopically. Many lines of inquiry have provided evidence in support of this model: clusters of enhancers and promoters are reported to intersect well with TADs (Shen *et al.*, 2012), co-regulated genes tend to share the same TAD (Symmons *et al.*, 2014), hypermethylation of CTCF binding sites results in reduced CTCF binding, reduced interaction insulation and subsequent ectopic transcription in cancer (Flavahan *et al.*, 2016), and lastly, an investigation into hierarchically nested domains shows transcriptional co-regulation to be maximal at the level of TADs (Zhan *et al.*, 2017). However, the relationship between TADs and transcriptional regulation is not straightforward. Although genes that share TADs tend to be co-regulated, this is not always the case (Dixon *et al.*, 2016), suggesting that additional mechanisms may operate in conjunction with TADs to precisely modulate transcriptional activity. This notion is supported by the observation that highly transcribed genomic domains contain housekeeping genes as well as tissue-specific genes (Versteeg *et al.*, 2003). Furthermore, the developmental gene cluster *Hoxd* is located directly at a TAD boundary and its genes are capable of interacting with enhancers in both neighbouring TADs (Rodríguez-Carballo *et al.*, 2017). This prompted Rodríguez-Carballo et al. to conclude that “[…] the physical separation of enhancers into two distinct TADs may not be a prerequisite for [them] to be properly operational in space and time”. This

suggests that certain regulatory interactions may form across TAD boundaries. Finally, upon perturbation of architectural proteins, transcriptional activity shows unexpectedly minor changes even though TAD organisation is affected. This prominent paradox illustrates that the influence of TAD organisation on transcriptional regulation is potentially more limited than initially thought. Upon various perturbations of cohesin, the reported number of transcriptionally affected genes ranges from 419 to 1984 (Schaaf *et al.*, 2009; Kagey *et al.*, 2010; Pauli *et al.*, 2010; Seitan *et al.*, 2013; Sofueva *et al.*, 2013; Rao *et al.*, 2017; Cuartero *et al.*, 2018). Moreover, upon rapid, targeted, near-complete depletion of the cohesin subunit RAD21 in human HCT-116 colon cancer cells, 87% (10,615) of all actively transcribed genes showed similar transcriptional activity (Rao *et al.*, 2017). The limited extent of misregulation suggests that TAD-directed promoter interactions are not the only mechanism directing transcriptional activity.

Collectively, TAD organisation appears involved in transcriptional regulation, likely by bringing promoters and enhancers into proximity. However, it is unlikely that TADs are the sole mechanism that governs promoter enhancer pairing. It remains to be seen to what degree additional mechanisms operate independently from TAD organisation.

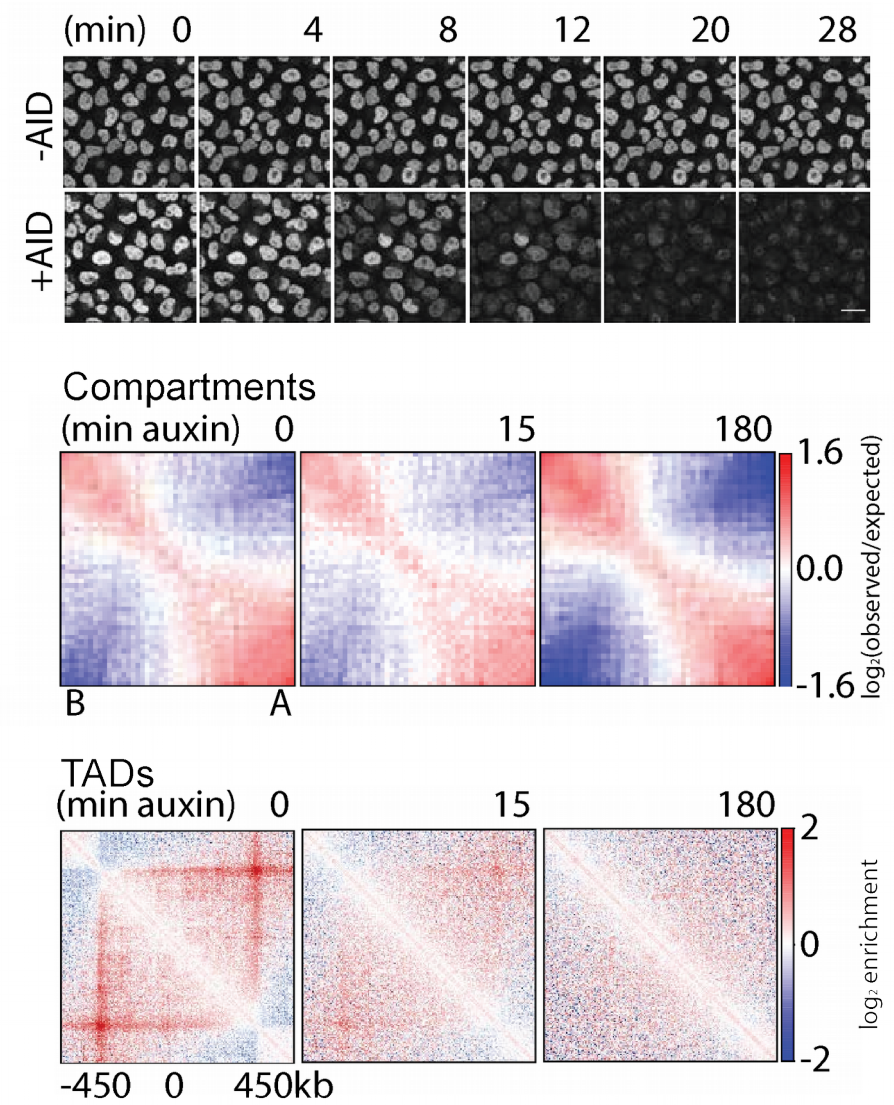

Figure 2. Cohesin perturbation by means of rapid depletion of SCC1 results in a strong reduction of TAD organisation, while compartmentalisation is preserved. Figure adapted from (Wutz *et al.*, 2017). The top row shows the decrease in fluorescence from SCC1-IAA1771-GFP upon activation of the AID system. The middle row shows preservation of compartmentalisation while the bottom row shows loss of TAD organisation.

1.1.5 Promoter-enhancer interactions

In previous sections, I alluded to transcriptional regulation through promoter-enhancer interactions. Here I define promoter-enhancer interactions in the context of spatial organisation. For a description of the components of transcription and its regulation, see chapter 1.2.

Long-range interactions between gene promoters and distal regulatory elements are a well-established mechanism of transcriptional control. The activity of transcriptional machinery at the core promoter region is directed by a large array of regulatory proteins called transcription factors (TFs). In the initial assembly of transcriptional machinery on the DNA, general transcription factors (GTFs) bind to the promoter region and together with RNA polymerase II, form the preinitiation complex (see section 1.2.2). However, TFs are known to regulate transcriptional activity even when they are bound to distal regions, located up to mega bases from the promoter (Lettice *et al.*, 2002, 2003; Fulco *et al.*, 2016). To explain this phenomenon, two models have been proposed: scanning and looping. The first one stipulates that distally bound TFs slide along the DNA until they encounter a promoter whereas the second one involves the formation of a spatial structure where the intermittent DNA between a promoter and a distally bound TF forms a loop, allowing the TF to come in proximity with the promoter whilst remaining bound to the distal site (Blackwood *et al.*, 1998; Bulger *et al.*, 1999). It is important to note that neither the scanning model nor the looping model preclude transcriptional regulation by transient enhancer-promoter communication. I.e. transcriptional activity may be directed by briefly enabling TFs to relay regulatory signals. In the context of looping, this may occur directly through protein-protein interactions between enhancer-bound TFs and transcriptional machinery, or indirectly through interactions with the mediator complex (see section 1.2.3). In the context of scanning, TFs may periodically migrate from a regulatory element to a promoter. Since the switch between paused transcription and transcription elongation is governed by the preinitiation complex (see section 1.2.2) which in turn receives input from TFs, the rate of transmission of regulatory signals is considered a crucial step in the overall regulation of transcription.

Scanning may impact distinct promoter architectures differently, since “ubiquitous” and “developmentally regulated” promoters contain regulatory elements that are spread out over a larger genomic region than “adult” promoters (see section 1.2.1). Since scanning enables TFs to transmit transcription-regulatory effects over short genomic distances, this mechanism is likely more important for regulation of genes with a diffuse promoter architecture where regulatory elements are spread out over a genomic region. This indicates that scanning may be a relevant mechanism for

regulation of genes with “ubiquitous” or “developmentally regulated” promoters. However, it is unknown wheter scanning is capable of transmitting regulatory signals across genomic regions that impose a physical barrier, such as CTCF binding sites, TAD boundaries, PRC-bound regions or other promoters. In support of the looping model, mounting evidence implicates spatial genome organisation (including distal interactions) in transcriptional activity. Perhaps owing to this and to the shortcomings of the scanning model to explain regulation by distal elements, emphasis has shifted towards the looping model. The first direct indications that distal elements and genes come into physical proximity were found by two studies focusing on the β-globin locus (Carter *et al.*, 2002; Tolhuis *et al.*, 2002). Tolhuis et al. made use of chromatin conformation capture. They exploited the knowledge that genes in the β-globin locus are transcriptionally active in erythroid tissue but silent in brain tissue. In brain tissue the conformation appears linear, suggesting that the locus is not engaged in looping interactions. In erythroid cells however, the locus control region and the beta globin locus form a loop and come in close spatial proximity. Tolhuis et al. report that the locus control region was found to be DNAse I hypersensitive, suggesting that the chromatin is open and capable of binding TFs. Furthermore, the DNA that was “looped out” contains inactive beta globin genes, illustrating how a distal regulatory element can skip over genes to regulate more than just its two neighbours (in contrast to the scanning model). Carter et al. use a biotin labelling system that marks all proteins in the vicinity of an actively transcribed gene. If a distal element is located in close spatial proximity to a promoter, the proteins bound to it are labelled with biotin. Subsequent streptavidin purification allows for analysis of the regions involved. Carter et al. used this system to investigate the regulation of the β-globin gene *Hbb-b1*, expressed in mouse livers, and were able to show that the enhancer element HS2 was significantly more labelled with biotin than control regions. This showed that HS2 was in close spatial proximity to the promoter of *Hbb-b1*, even though the two are ~50kbp apart. Taken together, these two experiments showed that promoters and distal elements can indeed interact through looping in addition to suggesting that this is involved in transcriptional regulation.

A separate line of investigation validates the notion that transcriptional regulation is supported by distal enhancers, by showing that dysfunctional distal regulatory elements can explain a range of developmental disorders. Lettice et al. showed that

the gene *Shh* is misregulated when a distal regulatory region is disrupted, leading to polydactyly (Lettice *et al.*, 2002, 2003). The regulatory region is ~800kbp away from the *Shh* gene, in addition to being located within an intron of the functionally unrelated gene *Lmbr1*. Another example is the developmental condition Pierre Robin sequence (a type of cleft palate) that was shown to be related to several genetic variants (Benko *et al.*, 2009) that implicate a highly conserved distal regulatory element of *Sox9* located ~1.44Mb upstream of this gene. Additionally, Benko et al. show presence of CTCF, the enhancer mark H3K4me1, and the transcriptional coactivator P300 at this distal element, which supports the notion that it is an enhancer (see section 1.2.4.2). In a separate investigation, Benko et al. show phenotypic evidence of functional misregulation of *Sox9*. Since *Sox9* is a sex determination gene, Benko et al observe sex reversal when a distal region that is located ~0.5Mb upstream of *Sox9* is disturbed (Benko *et al.*, 2011). Later research revealed that such sex reversal can be caused by an intra-TAD genome duplication that may disrupt communication between the *Sox9* promoter and enhancers, while the enveloping TAD appears preserved (Franke *et al.*, 2016). The latter suggests that sub-TAD structures may influence transcriptional activity. Together, these findings on *Sox9* highlight the fact that correct functioning of distal regulatory regions is paramount for proper development. Similarly, Uslu et al. show that the *Myc* transcription factor is related to facial development and that cleft lip or cleft palate malformations can occur upon its misregulation (Uslu *et al.*, 2014). They identify a distal region that regulates *Myc* and show that it has enhancer characteristics marked by H3K4me1 and H3K27ac. The enhancer region is located ~1.2 Mb downstream of *Myc*. Finally, Lupiáñez et al. reengineered naturally occurring human limb malformations in mice by disrupting TAD organisation, leading to ectopic promoter interaction wiring and subsequent transcriptional misregulation of *Wnt6*, *Ihh* and *Pax3* (Lupiáñez *et al.*, 2015). Taken together, evidence from these studies suggests that distal regions influence transcriptional activity and variants within these regions underpin several well studied developmental disorders. Lastly, there are indications that TAD organisation contributes to correct promoter interaction wiring, although it is unclear if this constitutes a common mechanism for all genes and to what extent promoter interactions may form independently of TADs.

In recent years, unbiased, high-resolution, genome-wide chromosomal interaction analyses have provided evidence that long-distance promoter interactions occur throughout the genome (Jäger *et al.*, 2015; Schoenfelder *et al.*, 2015; Javierre *et al.*, 2016). Furthermore, it has been suggested that multiple enhancer elements may cluster spatially and that these clusters are involved in tissue specific transcriptional regulation (Elizabeth Ing-Simmons *et al.*, 2015). Conversely, it has been shown that a single enhancer can simultaneously regulate two genes (Fukaya *et al.*, 2016). However, the majority of promoter interacting-regions (PIRs) appear to lack enhancer elements and may represent a structural rather than a functional class of interactions (Schoenfelder *et al.*, 2019). Together, these findings highlight that promoter interactions represent a fundamental level of genome organisation and transcriptional regulation.

The mechanisms that tether promoter interactions and govern promoter-enhancer specificity remain a subject of continued investigation. Although TADs have been suggested to organise DNA spatially in a way that facilitates promoter-enhancer proximity and supports interactions, it is likely that TADs alone are insufficient to explain all promoter interaction wiring. Additional mechanisms may function alongside TAD organisation to provide further regulatory complexity and specificity. For instance, the mediator complex has been suggested to connect promoters and enhancers, in addition to interacting with cohesin (Kagey *et al.*, 2010). The mediator complex is involved in integrating and transmitting TF regulatory signals to RNA polymerase II and the pre-initiation complex at promoter regions (Malik *et al.*, 2010). This large complex consists of 26 core proteins (although the composition may vary owing to its modular nature) and it is implicated in connecting promoters and enhancers (Malik *et al.*, 2010; Allen *et al.*, 2015; Soutourina, 2018). Another potential factor mediating TAD-independent physical promoter-enhancer tethering is the widely expressed zinc finger TF Yin Yang 1 (YY1). Recent analyses suggest that DNA-bound dimers of this protein are capable of supporting promoter enhancer interactions and it has been suggested as a candidate for halting loop extrusion (Beagan *et al.*, 2017; Weintraub *et al.*, 2017). Although YY1 has previously been identified as a cofactor of CTCF, in the context of allele-specific transcriptional regulation (Donohoe *et al.*, 2007), ~70% of YY1-bound sites are not co-bound by CTCF (Schwalie *et al.*, 2013). Furthermore, YY1 appears to be stabilised through

interactions with eRNAs (non-translated RNA transcripts of enhancers. See section 1.2.4.2), suggesting that YY1 can stably bind to DNA independently of CTCF (Sigova *et al.*, 2015).

Collectively, a number of mechanisms have been suggested that may bring promoters and enhancers in physical proximity to functionally interact and regulate transcription. Although TADs are likely involved in this process, other mechanisms may operate simultaneously. Therefore, it is important that the precise interplay between promoter interaction and higher order spatial organisation are integrated and studied in an unbiased, genome-wide way.

1.2 Transcriptional regulation

1.2.1 Genes and the core promoter region

Genes form the template for the synthesis of all proteins in an organism, encoding the information that instructs cellular machinery to string sequences of amino acids together in precise order, such that functional proteins are produced. The human genome contains roughly 20,000 to 25,000 (International Human Genome Sequencing Consortium, 2004). In the process of transcription, RNA Polymerase "reads" a gene and produces an intermediate messenger RNA (mRNA) which forms the template for the aforementioned amino acid polymer. Although the idea of a genetic carrier of hereditary information is not new, the precise definition has changed with increased understanding of the genome (Gerstein *et al.*, 2007). Gerstein et al. proposed the following formal universal definition of a gene, which encompasses three aspects:

1. "A gene is a genomic sequence (DNA or RNA) directly encoding functional product molecules, either RNA or protein"
2. "In the case that there are several functional products sharing overlapping regions, one takes the union of all overlapping genomic sequences coding for them"
3. "This union must be coherent—i.e., done separately for final protein and RNA products—but does not require that all products necessarily share a common subsequence"

Genes encode proteins or non-coding RNAs (ncRNA) as functional products. However, parts of the genes known as introns are not found in the mature RNAs generated on their template, and are instead spliced out of them. Alternative splicing can give rise to multiple gene products that are considered to be variants of the same gene. A single gene therefore can encode a range of morphologically and functionally distinct proteins. In short, Gerstein et al. summarize the definition as: “The gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products”.

The core promoter region (henceforth: promoter) is of fundamental importance for transcription initiation. It attracts the proteins that facilitate transcription, the pre-initiation complex (PIC), and it governs the position and intensity of transcription initiation (Kadonaga, 2012). Promoters are regions on the genome that contain DNA elements with sequence-specific motifs. These elements are protein-binding sites which attract distinct proteins and which can be present in varying numbers and at variable distances from the transcription start sites (TSSs). A gene can have one or more TSSs and the core promoter region is located in their direct vicinity. The location of promoter elements is usually indicated relative to the TSS and they may be located downstream of the TSS. As an example, the TATA box and the upstream transcription factor IIB (TFIIB) recognition element (BREu) are among the most well-known human promoter elements. Both elements are located upstream of the TSS (-24 to -31 bp, and -32 to -38 bp), and they associate with the PIC components TATA-box-binding protein (TBP) and TFIIB respectively (Haberle *et al.*, 2018). Promoter composition can vary wildly from gene to gene. This is illustrated by the fact that the strongly conserved TATA box is found in only a minority of promoters (Kadonaga, 2012; Haberle *et al.*, 2018). Additionally, promoter elements have been identified that are located at different distances from the TSS for different genes (Haberle *et al.*, 2018). Broadly, promoters can be grouped into three categories based on their composition, gene function and chromatin properties: adult, ubiquitous, and developmentally regulated (Lenhard *et al.*, 2012):

- ‘Adult’ promoters are involved in transcriptional regulation of genes that are active in fully differentiated cells. They tend to contain few TSSs, nucleosome positioning in their promoter regions is typically imprecise, they frequently

contain TATA box and initiator motifs, and they often bear the canonically active epigenetic marks H3K4me3 and H3k27ac

- ‘Ubiquitous’ promoters correspond to genes that are generally active throughout the cell cycle. They tend to contain many TSSs, in addition to individual CpG islands. They are typically free of nucleosomes, although nucleosomes are precisely positioned toward the edges of the promoter. Furthermore, ubiquitous promoters commonly bear the active epigenetic marks H3K4me3 and H3k27ac
- ‘Developmentally regulated’ promoters belong to genes that are precisely regulated during development and differentiation. They tend to contain many TSSs, in addition to many CpG islands and they are frequently epigenetically poised: bearing the active mark H3K4me3 simultaneously with the repressive PRC mark H3K27me3. Finally, developmentally regulated promoters are more likely to interact with multiple distal enhancers

Classification of promoters into these three categories is not straight-forward since many determinants are variable. However, current estimates associate ~70% of all promoters with CpG islands, suggesting that the majority of all promoters fall into the categories ‘ubiquitous’ or ‘developmentally regulated’ (Haberle *et al.*, 2016).

Collectively, genes are well-defined entities in the genome that can be transcribed to ultimately produce proteins and ncRNAs. The core promoter region serves as the starting point for transcription and the various identified promoter architectures related to the functions of the protein that the gene encodes.

1.2.2 The preinitiation complex

Transcription is performed by the protein complex RNA polymerase. Bacteria typically rely on one type of RNA polymerase whereas eukaryotes possess three kinds: RNA polymerase I, II and III (Roeder *et al.*, 1969, 1970; Ebright, 2000). The vast majority of genes are transcribed by Pol II, whereas RNA polymerase I transcribes ribosomal RNAs and polymerase III transcribes transfer RNAs and small nucleolar RNAs (Weinmann *et al.*, 1974; Warner, 1999). Eukaryotic transcription by Pol II is supported by a range of proteins that bind at the core promoter to form the preinitiation complex (PIC). In this complex, RNA polymerase II (Pol II) is responsible

for gene transcription to form the mRNA intermediate. The leading model for PIC assembly involves six GTFs that recognize promoter elements and that bind to it in a sequential order: first, TFIID binds to the promoter region where it can recognise TATA box or Inr elements. This is followed by TFIIA, TFIIB, Pol II in complex with TFIIF, TFIIE and finally TFIIH (Thomas *et al.*, 2006). Note that Pol II cannot bind to DNA without the PIC.

There is broad agreement that TFIID recruitment plays a crucial role as the first step in PIC assembly and therefore that TFIID forms the first step at which transcription can be regulated (Thomas *et al.*, 2006; Haberle *et al.*, 2018). TFIID consists of multiple subunits, distinct compositions of which are known to associate preferentially with specific types of promoters (e.g. ‘ubiquitous’ versus ‘adult’). Furthermore, TFIID recruitment to the promoter is likely stabilized by the enhancer-bound activator protein Rap1 which can interact with TFIID through formation of a DNA loop (Papai *et al.*, 2010). These examples illustrate that TFIID can serve as a point of transcriptional control in differing ways. Upon completion of PIC assembly, a ~15bp stretch of DNA is unwound at the TSS by TFIIH, which allows a single strand of DNA to enter the active site of Pol II, followed by initial synthesis of ~30bp of RNA (Grünberg *et al.*, 2013). Subsequently, Pol II may either pause or detach from the PIC to continue transcription elongation. The switch between pausing and elongation is thought to be related to phosphorylation of the Pol II carboxyl-terminal domain (CTD) (Phatnani *et al.*, 2006) and it considered a crucial step in the overall regulation of gene expression. This is illustrated by the observation that developmentally regulated genes in embryonic stem cells frequently show Pol II pausing at the core promoter region (Guenther *et al.*, 2007). In *D. melanogaster* 30% of all genes reportedly show both pausing and transcription, suggesting that a pausing rate may fine-tune mRNA synthesis (Adelman *et al.*, 2012).

Taken together, transcription initiation is supported by a range of proteins that collectively form the PIC. Various forms, modifications and interactions may influence the PIC such that transcription occurs at varying rates.

1.2.3 Regulation of transcriptional activity by transcription factors

There are many points at which transcription can be regulated. Broadly, chromatin can be modified, PIC recruitment and assembly can be modulated, and promoter proximal pausing can be regulated to increase the number of actively transcribing Pol II complexes on a gene. To coordinate transcription in a gene specific manner, TFs can recruit chromatin remodelling enzymes to modify chromatin compaction and nucleosome positioning at the promoter region (Agalioti *et al.*, 2000; Lomvardas *et al.*, 2002; Berger, 2007). This can increase accessibility for PIC components as well as for additional TFs, conveying their own regulatory signals (Li *et al.*, 2007). TFs can bind to distinct regulatory DNA regions, often containing conserved DNA motifs, to exert regulatory influence on target promoters (Vaquerizas *et al.*, 2009; Spitz *et al.*, 2012). Binding specificity between TFs and DNA-regulatory elements arises through complementarity between the protein surface of the TF and the nucleotide composition at the DNA element, which allows TFs to recognise and bind to distinct elements while ignoring others (Todeschini *et al.*, 2014). TFs can influence the sequential assembly of the PIC at many points to increase or decrease the rate of PIC formation (Thomas *et al.*, 2006). Additionally, TFs can interact directly with the PIC or Pol II, influencing the stability of the PIC or triggering the release of paused Pol II (Adelman *et al.*, 2012). However, direct contact between TFs and the PIC is not required; the mediator complex can relay TF signals, providing an indirect mechanism for transmission and integration of regulatory signals (Malik *et al.*, 2010; Allen *et al.*, 2015; Soutourina, 2018). In humans, mediator consists of 26 core subunits (Allen *et al.*, 2015) which collectively act as an interface between TFs and the PIC, allowing multiple TFs to bind and integrating their regulatory signals (Poss *et al.*, 2013). It has been suggested that the conformational changes that the mediator complex undergoes upon interaction with TFs, mechanistically convey regulatory signals to the PIC, regulating release of paused Pol II (Taatjes *et al.*, 2004; Meyer *et al.*, 2010; Poss *et al.*, 2013).

Collectively, TFs direct transcriptional activity by influencing the promoter region directly or by imposing regulatory signals on the PIC, supported by the mediator complex. Chromatin accessibility, PIC assembly and Pol II pausing are the focal points of transcriptional regulation. The fact that TFs can indirectly communicate

regulatory signals to the PIC (through the mediator complex), enables TFs at distal enhancers to control transcriptional activity. A promoter may be regulated by several enhancers which may be located kilobases or even more than a megabase in either direction of the promoter (Benko *et al.*, 2009; Uslu *et al.*, 2014; Rodríguez-Carballo *et al.*, 2017). In the next section, I provide an overview of distal regulatory elements.

1.2.4 DNA regulatory elements

Various classes of regulatory elements have been identified, among which the previously detailed core promoter region is a prominent example. It is unknown what proportion of the genome encodes functional elements, although comparative genomics estimates range from 3% to 8% of all base-pairs, including genes (The ENCODE Project Consortium, 2011). However, these estimates rely on sequence conservation driven by selective pressure. Such approaches disregard species-specific genetic elements as well as evolutionary neutral functional elements (i.e. loss of a regulatory element and subsequent loss of target gene transcriptional activity may not significantly affect fitness), as well as those functional elements that are preserved through evolutionary turnover rather than direct conservation. Furthermore, such approaches may fail to detect conservation of very short functional elements, fragmented elements, or elements that are located in repetitive regions of the genome. Together, these limitations of comparative genomics approaches suggest that an even larger proportion of the genome contains functional elements. In this section I describe four canonical classes of regulatory elements, followed by a description of current strategies in novel regulatory element detection.

1.2.4.1 Promoter proximal elements

Promoter-proximal elements can be found in the order of hundreds of base pairs upstream of TSSs. These harbour TF binding sites which can attract regulatory proteins directly to the vicinity of the promoter (Maston *et al.*, 2006). Additionally, unmethylated CpG islands are often found proximally to promoters and have historically been used to indicate the presence of a gene (Bird, 1987). This class of regulatory element is closely tied to promoter organisation and are sometimes referred to as proximal enhancers, in contrast to those that are located at greater distances from the promoter region.

1.2.4.2 Enhancers

The first enhancer was discovered almost four decades ago when it was shown that a 72bp DNA sequence from the SV40 virus could increase transcription of a rabbit beta globin reporter gene in HeLa cells (Banerji *et al.*, 1981; P.Moreau *et al.*, 1981). Importantly, it was noted that the enhancer sequence functioned at variable distances from the reporter gene and that the enhancer was not dependent on orientation. Since then, it has become apparent that the human genome contains many thousands of enhancer sequences that operate collectively to precisely regulate transcriptional activity according to cell-type or developmental stage (Long *et al.*, 2016). Enhancers play a crucial role in guiding transcriptional regulation in accordance with developmental programmes and dysfunctional enhancers can cause developmental abnormalities (see section 1.2.4.2). Enhancers typically consist of groups of 6-10bp long TF binding sites, bearing distinct sequence motifs (Bulger *et al.*, 2011; Shlyueva *et al.*, 2014), and are canonically marked by H3K4me1 in addition to the H3K27ac mark (International Human Genome Sequencing Consortium, 2004; Heintzman *et al.*, 2009). Recently, a novel class of non-canonical enhancers has been identified in *Mus musculus* embryonic stem cells which bear H3K122ac while lacking H3K27ac (Pradeepa *et al.*, 2016). Additional features of enhancers are: chromatin accessibility, Pol II and TF binding and association with the mediator complex (Shlyueva *et al.*, 2014). Furthermore, enhancer elements have been shown to be transcribed, leading to non-coding eRNA (Li *et al.*, 2016). Extended regions of the genome with multiple enhancer elements that associate strongly with the mediator complex are sometimes distinctly referred to as super-enhancers (Whyte *et al.*, 2013; Elizabeth Ing-Simmons *et al.*, 2015; Schoenfelder *et al.*, 2019). Promoters themselves may also exhibit distal enhancer-like characteristics where a distal promoter is regulated by a separate promoter of a different gene (Li *et al.*, 2012; Dao *et al.*, 2018). Searches for novel enhancer elements often rely on detection of one or more of these properties (Visel, Rubin, *et al.*, 2009). Interestingly, a novel class of regulatory element has recently been described that appears to lack any known epigenetic markers (Rajagopal *et al.*, 2016). This apparent absense of epigenetic marks suggests that the marks themselves do not convey enhancer activity.

1.2.4.3 Insulators

Insulators are DNA elements that can limit the effects of neighbouring elements by forming a barrier to their influence. In the seminal work on insulators and boundary elements, Kellum and Schedl describe the capacity of “specialized chromatin structures” that prevent the regulatory interaction between the promoter hsp70 and the enhancer yp-1 in *D. melanogaster* (Kellum *et al.*, 1991, 1992). One of the most prominent vertebrate insulators is the CTCF binding site that contains a distinct sequence motif that is recognised by zinc finger domains of CTCF proteins (Ong *et al.*, 2014). Computational analyses of sequence conservation among species suggest ~14,000 CTCF binding sites in vertebrate genomes (Wallace *et al.*, 2007) although this is likely an underestimate due to the aforementioned limitations of sequence conservation driven approaches. When bound to DNA, CTCF can prevent enhancers from exerting regulatory influence on neighbouring genes (Bell *et al.*, 1999). CTCF has been suggested to halt the spread of heterochromatin (Narendra *et al.*, 2015), although near-complete depletion of CTCF has been shown not to induce H3K27me3 spreading (Nora *et al.*, 2017). CTCF binding sites provide genomic locations that delimit DNA accessibility and narrow the space of regulatory interactions, which has been shown to relate to developmental processes (Wallace *et al.*, 2007; Narendra *et al.*, 2015). The role that CTCF plays in spatial organisation through loop formation (see section 1.1.5) forms a likely mechanism by which insulation occurs. It is likely that loop extrusion supports CTCF-based insulation although indications that CTCF may interact with other proteins to form loops (Yusufzai *et al.*, 2004) suggest additional mechanisms. One such protein is YY1 (Beagan *et al.*, 2017). The fact that YY1 is known to be capable of supporting DNA loops, reinforces the notion that looping is a major mechanism for insulating DNA interactions (Beagan *et al.*, 2017; Weintraub *et al.*, 2017). Taken together, insulator elements shape the DNA interaction space by preventing interactions from crossing them, in addition to directing compaction.

1.2.4.4 Silencers

Silencers confer transcriptionally repressive effects on transcriptional activity. Interestingly, silencers share characteristic properties of enhancers such as protein binding specificity and distal regulatory influence (Ogbourne *et al.*, 1998). However,

in recent years the term ‘silencer’ has become less common, perhaps owing to their similarity to enhancers. The notion that a silencer is simply an enhancer that binds repressive proteins is illustrated by the fact that repressive Polycomb response elements are not defined by a conserved sequence but rather, contain multiple distinct sequence motifs that are known binding sites for TFs (Schwartz *et al.*, 2007). Furthermore, it has been shown that YY1 can stimulate, as well as repress transcriptional activity (Shrivastava *et al.*, 1994), highlighting that TF activity versus repressor activity can be dependent on the protein and its modifications/interactions rather than the DNA element it binds to. Taken together, it appears that the concept of silencer elements as distinct entities is gradually shifting towards the notion that enhancers may convey repressive signals, rather than being limited to conveying stimulating signals only. This section was included to highlight why silencers were not studied in the context of promoter interactions and promoter interaction rewiring in this thesis.

1.3 Current methods in the study of spatial genome organisation

1.3.1 Microscopy

Historically, spatial genome organisation has been studied using microscopy and fluorescent in situ hybridization (FISH). FISH allows labelling of target genomic loci with fluorescent dyes which can be observed with a light microscope. FISH uses oligonucleotides (oligos) to target genomic loci and typically used to be limited to detection of several genomic loci, owing to the high cost of oligo synthesis. In recent years this cost has fallen dramatically, enabling approaches such as Oligopaint (Beliveau *et al.*, 2012), that rely on the creation of large and complex libraries of oligonucleotides, to target multitudes of genomic loci simultaneously. Furthermore, super-resolution microscopy techniques such as STORM (Rust *et al.*, 2006) and PALM (Betzig *et al.*, 2006) employ photoswitching/photoactivation to resolve smaller structures than ever before. In a recent single cell imaging approach, a combination of STORM and Oligopaint was used to obtain a genomic resolution of 30kb, enabling visualisation of sub-TAD organisation (Bintu *et al.*, 2018). In this approach, Oligopaint probes were adapted to label 30kb segments of the genome with distinct 20 nucleotide long readout sequences. These were then labelled with dyed probes,

complementary to the readout sequences. Sequential labelling with readout probes then provided snapshots of consecutive 30kb segments, which were used in three-dimensional STORM and three-dimensional imaging to create three- dimensional images within single cells. Although microscopy-based methodologies have shown impressive progress, they are still limited in terms of the number of genetic loci that can be analysed simultaneously. In other words, currently they cannot be applied in a genome-wide fashion. However, continuous improvements in this field will likely expand the current capabilities and offer a valuable approach to spatial genome analyses, complementary to sequencing-based methodologies.

1.3.2 Chromatin conformation capture and derivatives

Chromatin conformation capture (3C) based methods rely on DNA fragmentation and re-ligation to map three-dimensional organisation onto a one dimensional reference genome (Dekker *et al.*, 2013; Schmitt *et al.*, 2016). In most 3C-based analyses, chromatin is cross-linked using formaldehyde to covalently lock the spatial organisation of the nucleus into place. Subsequently, DNA is fragmented, usually by means of restriction digestion, followed by re-ligation. This step allows spatially neighbouring fragments ligate, even though they may be located distally on the linear genome sequence. The ligation frequency is assumed to reflect the DNA interaction frequency and is used to infer spatial proximity. 3C detects single ligation products by polymerase chain reaction (PCR) amplification using locus-specific primers (Dekker *et al.*, 2002). Several 3C derivative technologies have since been developed to overcome the limitation of 3C to pairs of loci (one versus one), thereby extending the capabilities to: one locus versus many (4C) (Splinter *et al.*, 2012), many loci versus many loci (5C) (Dostie *et al.*, 2006) and all versus all loci (Hi-C) (Lieberman-Aiden *et al.*, 2009). The latter includes a step where biotinylated nucleotides are integrated after fragmentation and before ligation. The resulting biotinylated ligation junctions correspond to ligation events between pairs of fragments. Using streptavidin-based purification, a genome-wide library of spatially proximal DNA pairs is constructed, which is then mapped onto a reference genome after sequencing. Although this enables detection of all versus all loci, the attainable resolution is limited by protocol efficiency (e.g. ligation and fill-in efficiency) and the sequencing depth. Since the combinatorial space of all versus all DNA fragments is vast; to interrogate all sequence pairs requires a sequencing depth that is inefficient in terms

of the amount of data that is produced and may be costly. E.g. at a commonly employed sequencing depth on the order of 10^8 read pairs, the resulting Hi-C contact map will be at 100kb resolution. In other words, sequenced fragment counts are binned into 100kb wide bins. To illustrate, such a resolution is insufficient for TAD detection.

To overcome these limitations, several *C-variants have been developed which select specific fragments, thereby enriching a fragment library for DNA interactions of interest while reducing unwanted and noisy interactions. Both the Chromatin Immunoprecipitation Interaction Assay with Paired End Tagging (ChIA-PET) and HiChIP select DNA fragments that interact with target proteins (Fullwood *et al.*, 2009; Mumbach *et al.*, 2016). For example, this enables the construction of a DNA interaction library where the fragments interact with Pol II, thereby enriching for promoter-enhancer interactions (Li *et al.*, 2012). However, reliance on immunoprecipitation comes with its drawbacks: a reliable antibody must exist for a target protein, antibody sensitivity to allosteric variants needs to be well characterised and finally, when comparing data between conditions, it is inherently problematic to distinguish between differences in DNA interaction frequency, protein binding, and immunoprecipitation efficiency. To overcome these challenges methodology, Capture Hi-C (CHiC), instead uses sequence capture of Hi-C libraries to unbiasedly enrich for fragments with a distinct nucleotide composition (Hughes *et al.*, 2014; Mifsud *et al.*, 2015; Schoenfelder *et al.*, 2015). The sequence capture step uses complementary probes to recognise and bind a set of target fragments. Promoter CHiC (PCHiC) uses a set of probes to target nearly all promoters simultaneously, leading to a roughly tenfold enrichment in read-pairs involving promoter elements when compared to Hi-C (Mifsud *et al.*, 2015; Schoenfelder *et al.*, 2015). The combination of unbiased enrichment for promoter fragments and the enrichment of such interactions compared to conventional Hi-C, makes PCHiC an excellent technique for genome-wide analysis of promoter interactions.

1.3.3 Hi-C: TAD detection

Upon discovery of TADs, Dixon et al. developed an algorithm to detect them (Dixon *et al.*, 2012). This algorithm utilises the directionality index (DI), which is a metric that describes the directional bias of DNA interactions in a contact matrix. It utilises a sliding window approach where the number of interactions that map from the window

to an upstream interval and the number of interactions that map from the window to a downstream interval are compared to the expected number of interactions under the null hypothesis. As such, the expected number of interactions E, is the mean number of interactions between A and B. The upstream interval A, the downstream interval B and E are then used to calculate the DI as follows:

$$DI = \left(\frac{B-A}{|B-A|} \right) \left(\frac{(A-E)^2}{E} + \frac{(B-E)^2}{E} \right)$$

The first component of this formula imposes a change of sign when A≠B. The second component of the formula compares the observed numbers of reads to the expected number under the null hypothesis. This results in a negative DI when the upstream window A shows more interactions than the downstream window B. Conversely, the DI is positive when downstream window B shows more interactions than the upstream window A. Furthermore, the DI increases in amplitude when the total number of interactions increases. The resulting DI values per window (bin) may then be processed further using a three-state hidden Markov model or kernel density smoothing. A TAD is then defined in an interval where the DI is positive, continues through the consecutive change of sign from positive to negative, and ends where the DI changes sign from negative to positive or to zero.

Various additional approaches to TAD detection have been described, such as bin-aggregation into hierarchical domains (Weinreb *et al.*, 2016; Zhan *et al.*, 2017), DNA interaction network analysis (Yan *et al.*, 2017; Norton *et al.*, 2018), or use of an alternative metric such as the insulation score (Sofueva *et al*., 2013) (Crane *et al.*, 2015). However, benchmarking TAD detection tools is problematic since there is little consensus on metrics for quality control and because no “gold standard” TAD partitioning exists. In fact, there is considerable evidence that a single optimal TAD partitioning may not be sensible at all. E.g. evidence exists that TADs can be dynamic in bulk detection (Narendra *et al.*, 2016; Chathoth *et al.*, 2019) as well as in single cell analysis (Boettiger *et al.*, 2016), highlighting that TADs may not represent a stable partitioning of the genome, but rather, define intervals where domain organisation is likely to be present. Notably, the cell cycle has been suggested as a source of deterministic variance in TAD organisation (Nagano *et al.*, 2017), indicating

that controlling for cell cycle heterogeneity may improve TAD detection and its reproducibility.

1.3.4 CHiC: interaction detection

Conventional Hi-C signal detection algorithms cannot analyse the non-square interaction matrices that are the result of asymmetric C-technologies such as CHiC and HiChIP. To robustly detect interactions in CHiC data, the CHiCAGO algorithm (Cairns *et al.*, 2015) was specifically developed for CHiC data analysis. In contrast to existing Hi-C analysis algorithms, CHiCAGO was designed to address the fact that noisy background counts are generated by two distinct processes: distance dependent Brownian collisions and technical variability. CHiCAGO models the expected interaction frequency using the Delaporte distribution, where the negative binomial distribution is used to model discrete interaction frequencies according to Brownian motion and the Poisson distribution is used to model the technical noise. Interactions are deemed significant when the interacting regions show a read coverage that exceeds the expected Delaporte background level in a one-tailed hypothesis test.

1.4 Aims

My thesis aims to examine two key issues in the interplay between TADs, promoter interactions and transcription. First, to investigate to what extent promoter interactions are insulated by TAD boundaries. Second, to investigate the response of promoter interactions and transcription to perturbations of architectural proteins.

2 Materials and Methods

All analyses were carried out in the R statistical environment unless stated otherwise. Extensive use was made of the data.table package (https://CRAN.R-project.org/package=data.table) that extends data frames providing improved speed and memory efficiency.

2.1 Promoter capture Hi-C and Hi-C sequence alignment

Promoter capture Hi-C (PCHiC) data for eight human haematopoietic tissues (B-cells, CD4 T-cells, CD8 T-cells, Erythroblasts, Macrophages, Megakaryocytes, Monocytes and Neutrophils) were obtained from (Javierre *et al.*, 2016). These libraries were constructed using the *HindIII* restriction enzyme. Raw reads were processed using the Hi-C pipeline HICUP (Wingett *et al.*, 2015), which maps Hi-C reads (di-tags) against a reference genome (in our case, GRCh37), in addition to removing duplicate reads and filtering artefacts such as circularised and adjacent ligation events.

2.2 Hi-C data processing and TAD calling

Aligned Hi-C data were processed with HOMER (Heinz *et al.*, 2010). Hi-C reads were binned at a 40kb resolution and normalised using iterative correction (Imakaev *et al.*, 2012). Directionality index (DI) scores (Dixon *et al.*, 2012) were calculated using a 5kb step size, a 25kb window size and a 1Mb upstream and downstream window size, and were subsequently smoothed using 25kb kernel density smoothing. Topologically associating domain (TAD) boundaries were called as local extrema in DI transitions from negative to positive were detected. Two stringency settings were used to partition the genome into standard TADs and stringent TADs. Standard TADs were calculated by setting the DI score threshold (minIndex) to 0.5 and the DI score difference threshold between negative and positive extremes (minDelta) to 1. Stringent TADs were calculated by setting minIndex = 1 and minDelta = 2.

2.3 TAD merging

Replicate TAD partitionings were merged by using the fast overlap joins function foverlaps() from R package data.table. This function takes two tables of interval coordinates as input and returns pairs of intersecting and non-intersecting intervals. I calculated the Jaccard index for each pair of intersecting TAD intervals and

discarded those that were below a cut off of 0.75. I then merged the TAD intervals by taking the centre point between boundaries from replicates:

$$B_b = \frac{\left(R_b^1 + R_b^2\right)}{2}, \quad (1)$$

where B_b is the merged TAD boundary b and R_b^r is the boundary coordinate for replicate r of boundary b.

For TADs that show no intersection, the Jaccard score histograms (Figure 3) show the largest number of non-intersecting TADs from the two replicates (at Jaccard = 0).

2.4 PCHiC interaction calling

Promoter interactions were called using the CHiCAGO pipeline (version 1.1.5) (Cairns *et al.*, 2015). CHiCAGO models expected promoter interaction frequencies based on the Delaporte distribution, which has a negative binomial component and a Poisson component. The negative binomial component is used to model discrete interaction frequencies according to Brownian motion, whereas the Poisson component is used to model technical noise. CHiCAGO corrects for multiple testing by means of a p-value weighting procedure based on the expected true positive rate at a given interaction distance estimated on the basis of consistency between biological replicates. CHiCAGO scores represent soft-thresholded –log weighted p-values. Since it was previously shown that enrichment of multiple chromatin marks is maximised at promoter interacting regions (PIRs) at a CHiCAGO score of 5 (Cairns *et al.*, 2015), this score was used to detect significant promoter interactions.

2.5 Clustering CHiCAGO scores

K-means clustering was used to partition promoter interaction scores across samples. The Hartigan-Wong clustering algorithm implemented in the function kmeans()from the R package stats was with a maximum of 1000 iterations and 25 initial random configurations. Prior to clustering, CHiCAGO scores were asinh-transformed and capped at median + 3MAD. The number of clusters k was determined by iteratively performing k-means clustering with k-values ranging from 2 to 16 and choosing a partitioning with a low k as well as a low total within-cluster sum of squares value. Additionally, partitionings that resulted in multiple clusters with similar centroid values were avoided. Using these criteria, k = 5 and k = 13 were

selected for the cell-cycle samples and the architectural protein perturbation samples, respectively (see Figure 13 and igure 29).

2.6 Differential promoter interaction calling (Chicdiff)

Chicdiff (version 0.2) was used to detect differential promoter interactions between SCC1-AID control (Aux-) and depleted (Aux+) samples, and between CTCF-AID control (Aux-) and depleted (Aux+) samples. Chicdiff aggregates PIRs per bait (captured restriction fragment that contains a promoter) to account for "passenger interactions" and uses DESeq2 (Love *et al.*, 2014) with a custom scaling matrix to call differential promoter interactions based on the expected read counts per interaction computed by the Chicago package (Cairns *et al.*, 2015). Chicdiff then uses independent hypothesis weighting (Ignatiadis *et al.*, 2016) to account for promoter interaction distance effects. A Chicago score cut off of 5 was used to prioritise the potential "driver" differential PIRs within the aggregated pools. The normalisation procedure that Chicdiff uses by default has been modified since version 0.2, and the approach that was used here is available in current Chicdiff versions via the setting norm = "fullmean".

Promoter interactions were categorised as "lost" if the weighted p-value was ≤ 0.01 and the $\log_2$ fold-change was < 0 and the difference in CHiCAGO score ($SCC1_{control}$ - $SCC1_{depleted}$) was > 1. Promoter interactions were categorised as "gained" if the weighted p-value was ≤ 0.01 and the $\log_2$ fold-change was > 0 and the difference in CHiCAGO score ($SCC1_{depleted}$ - $SCC1_{control}$) was > 1. Promoter interactions were categorised as "maintained" if the weighted p-value was > 0.01 and the CHiCAGO score was >= 5 in both conditions and the absolute difference in CHiCAGO score (|$SCC1_{control}$ - $SCC1_{depleted}$|) was < 1.

2.7 Calculating the proportions of TAD boundary crossing promoter interactions

To calculate TAD boundary crossing (inter-TAD) promoter interaction frequencies, each restriction fragment (RF) was assigned to the most proximal TAD interval. Inter-TAD promoter interactions were defined as those whose baits (RF containing a gene promoter) intersected a TAD interval while the respective PIRs did not. The proportion of TAD-boundary crossing interactions in a given cell type was calculated by dividing the number of inter-TAD promoter interactions by the total number of promoter interactions where the bait intersects a TAD interval. To determine the

proportion of non-crossing (intra-TAD) interactions per promoter, the number of intra-TAD interactions of a given promoter is divided by the total number of interactions of that promoter. The expected TAD boundary crossing distribution was calculated by randomly reassigning baits to TADs, sampling from the set of baits that intersect a TAD. This randomization was performed 1000 times for each cell type and the mean +/- SD was reported.

2.8 Calculating promoter interaction frequencies with respect to TADs

To create diagrams showing the frequency of RFs with respect to TADs (Figure 16), the linear genomic distance between a given RF and the most proximal TAD boundary was normalised for the length of the TAD:

$$RF_n^{'} = \frac{\left(B_1 - RF_n\right)}{B_2 - B_1}, \quad (2)$$

where $RF_n^{'}$ is the normalised location of the nth (normalised with respect to the length of the most proximal TAD), B_1 and B_2 are the start and end boundary coordinates of the most proximal TAD, respectively, and RF_n is the centre coordinate of the nth restriction fragment. $RF_n^{'}$ is 0 when the RF is located at the TAD start boundary, and it is 1 when the RF is located at the TAD end boundary. $RF_n^{'}$ is calculated for each RF and the frequency distribution of RF_{1-N} is then calculated and visualised for the interval: $RF_n^{'} = [-2, 3]$. Note that this interval corresponds to the x-axis of the figures, showing the values [-2, -1, S, E, 1, 2]. The frequency distributions for multiple TAD partitionings are combined by calculating the mean and the standard error (SE) per bin.

For the “viewpoint window” analyses (Figure 17, Figure 38 and Figure 40), the same approach was taken, except that one of the interacting RFs (of a promoter interaction) was required to be located in an inter-TAD pentile interval, spanning 20% of the length of the TAD and combined for both flanks of the TAD, i.e.: proximal: 0-20% and 80-100%; intermediate: 20-40% and 60-80%; mid-TAD: 40-60%. The Peripheral and intermediate windows are combined by taking the reflection along the TAD-centre of the frequency distributions of the peripheral and intermediate windows on the “end” side of the TAD. The mean and SE are then calculated over the frequency distributions of the peripheral and intermediate windows on the “start” side

of the TAD and the reflected peripheral and intermediate windows on the “end” side of the TAD.

For the “half TAD+” analyses (Figure 22 and Figure 23), the viewpoint window analysis above was adapted such that the viewpoint window encompassed half the TAD, in addition to the region beyond the TAD boundary. I.e. this approach is not limited to analysing RFs within-TADs because it includes RFs located outside the TAD interval. This was followed by combining the frequency distributions on either side of the TAD centre.

2.9 TAD window enrichment analysis

Calculating RF enrichment in TAD windows (Figure 39) was performed by dividing the observed proportion of RFs in a pentile window (peripheral, intermediate and central) by the expected proportion according to the uniform distribution. For the central window the expected proportion is 0.2 (one fifth), while for peripheral windows, as well as for intermediate windows, the expected proportion is 0.4 (since there are two of each of these windows). Enrichment is then defined as:

$$Enr_w = \log_{10}\left(\frac{O_w / O_t}{E_w}\right) \quad (3)$$

Where Enr_w is the enrichment in window w (peripheral, intermediate or central), O_w is the observed number of interacting RFs in window w, O_t is the observed total number of interacting RFs in the TAD, and E_w is the expected proportion of RFs under the uniform distribution for window w (0.4, 0.4, or 0.2).

The enrichment scores are calculated for the four TAD partitionings (G1R1, G1R2, G2R1, G2R2) and the mean +/- 1SE are shown.

2.10 Public ChIP-Seq data sources and quality control

HeLa-specific ChIP-Seq data were obtained from two sources: the ENCODE project (Encode Consortium, 2013) and the Gene Expression omnibus (GEO) (Edgar, 2002).

ENCODE files were downloaded manually as BAM files, aligned against GRCh37. Quality control on ENCODE files was performed using the “Read coverage audits” QC metrics provided on the ENCODE website. Only entries with replicate data and

with a read length of ≥36 were included and files with severe or multiple read-depth or coverage limitations were excluded.

Unaligned data were downloaded from GEO as fastq files using the sra_fqdump command in Cluster Flow (Ewels *et al.*, 2017). Read quality was tested using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and FastQscreen (Wingett *et al.*, 2018) was used to test for contaminants and to ascertain that the sequencing material indeed originated from the human genome. Overall, 19 ChIP-Seq targets were excluded from further analysis (and were not included in the Supplementary Table). Subsequently, Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to remove adapters. Reads were aligned against GRCh37 using Bowtie2 (Langmead *et al.*, 2012)

2.11 Calculating scores per restriction fragment

2.11.1 ChIP-Seq

Read counts per RF were obtained by using htseq-count, requiring reads to have a minimum mapping score of 10 and with the –m flag set to 'union'. Subsequently, RFs that were shorter than 50bp or longer than 50kbp were excluded from further analysis. ChIP-Seq replicates were merged by taking the mean and rounding up to whole numbers. The Anscombe variance stabilising transformation was applied to the counts per RF (Harrison, 2015), with the dispersion parameter set to 0.4, followed by between-chromosome quantile normalisation separately for each ChIP-Seq target performed using the normalizeQuantile() function from the aroma.light R package (Bengtsson *et al.*, 2004).

OLS regression was then used to define the background relationship between the Anscombe-transformed, quantile-normalised scores per RF and the RF length. The independent variable was the RF length and the dependent variable was the transformed, normalised score. The studentised residuals from these regression models were taken as the ChIP-Seq scores per RF. Replicates were merged by taking the mean.

2.11.2 ATAC-seq

The ATAC-seq scores per RF were calculated identically to the ChIP-Seq approach, with four exceptions:

1. The read counts per RF were obtained by using the bamtobed() function from the BEDTools suite (Quinlan *et al.*, 2010). The ATAC-seq coordinates were then adjusted by +4 bp on the + strand and by -5 bp on the - strand to obtain the precise TN5 integration sites (Buenrostro *et al.*, 2013).
2. RFs with outlier scores were discarded (>99.99th percentile and <0.01st percentile), removing 326 RFs.
3. Additional quantile normalisation was applied to the studentised residuals. This normalisation step was applied between samples.
4. Signal strength bias between conditions was tested for by means of an MA-plot, showing negligible bias.

2.12 Integrating Ensembl regulatory build annotations with restriction fragments

HeLa-specific chromatin annotations from release 85 of the regulatory build were downloaded from the Ensembl FTP site:

ftp://ftp.ensembl.org/pub/grch37/release-85/regulation/homo_sapiens/

Using the foverlaps() function from the data.table R package, the regulatory build annotations were assigned to HindIII fragments (by genomic coordinates).

2.13 Regression analyses

2.13.1 LASSO logistic regression analyses

LASSO logistic regression used the rewiring category (*lost* or *maintained*) as the dependent variable. The independent variables were the studentised ChIP-Seq scores per PIR (mean over replicates per target), in addition to the CHiCAGO score and the $\log_{10}$(bp) interaction distance. Prior to regression, the continuous independent variables were scaled and centred such that mean = 0 and SD = 1. Since a PIR can have multiple interactions, the regressions were performed separately using the lower bound, mean and upper bound per PIR of the CHiCAGO score and interaction distance. PIRs showing an ambiguous rewiring response across the baits they contacted (i.e. *lost* as well as *maintained*) were excluded from this analysis.

LASSO logistic regression was performed using the glmnet() function from the glmnet R package (Friedman *et al.*, 2010), with the alpha flag set to 1 (for LASSO regression) and the thresh parameter set to 10^{-12}. Cross validation was performed

using the function cv.glmnet(). Significance was tested at a lambda score that minimises the number of parameters while staying within 1SE of the cross-validated error. The function fixedLassoInf() from the R package selectiveInference (Lee *et al.*, 2016) was used to infer p-values, with parameters tol.beta = 10^{-3}, and tol.kkt = 0.3.

TAD boundary crossing (inter-TAD-only) LASSO logistic regression was performed in the same manner, with the exception of the dependent variable, which in this case was a Boolean parameter that is set to 1 if a promoter interaction crosses a TAD boundary in all four TAD partitionings (G1R1, G1R2, G2R1, G2R2), and to 0 if a promoter interaction crosses no TAD boundary in all four TAD partitionings. Note that this excludes promoter interactions with ambiguous TAD boundary crossing.

2.13.2 Logistic regression: TAD boundary crossing versus the interaction distance and rewiring category of promoter interactions

Logistic regression was performed at the level of individual promoter interactions, where the Boolean variable representing consensus TAD boundary crossing defined as above was used as the dependent variable. The independent variables were the $\log_{10}$(bp) promoter interaction distance and the rewiring category (*lost*, *maintained* or *gained*). A 'sum to zero' contrast matrix was used. The R function allEffects() from the effects package (Fox, 2003) was used for visualisation.

2.13.3 Logistic regression: Cohesin and CTCF binding score at PIRs versus promoter interaction rewiring category

Logistic regression was performed on the level of individual promoter interactions, where the dependent variable was a Boolean that takes the value 1 when a promoter interaction is *maintained* and takes the value 0 when a promoter interaction is *lost*. The independent variables were the $\log_{10}$(bp) promoter interaction distance, the ChIP-Seq score per bait and the ChIP-Seq score per PIR. A 'sum to zero' contrast matrix was used. In addition to including the aforementioned independent variables, the interaction effect between the ChIP-Seq score at the bait and at the PIR were included in the model. The R function allEffects() from the effects package (Fox, 2003) was used for visualisation.

2.13.4 Ordinal logistic regression: promoter-enhancer interaction rewiring versus SLAM-seq signal for the respective gene

To select RFs containing active enhancers, ChIP-Seq scores per RF were used. RFs were marked active if the score for H3K4me1 was ≥95th percentile while simultaneously showing ≥95th percentile scores for H3K27ac *or* H3K4me3. RFs were marked not-active if the scores for all these three marks were <95th percentile.

Subsequently, the number of *lost*, *maintained* and *gained* PIRs containing active enhancers per bait were computed. Additionally, the number of non-active *lost*, *maintained* and *gained* PIRs per bait were computed. These numbers were then used to construct the independent variable defined as:

$$X = \frac{nGained - nLost}{nLost + nMaintained + nGained + 1} \quad (4)$$

where nGained, nMaintained and nLost represent the numbers of *gained, lost and maintained* promoter interactions per bait, respectively. The independent variables were computed separately on the active and the not-active tallies.

SLAM-seq data for the nascent transcriptional response upon SCC1 depletion was used to construct the dependent variable (see Figure 51). Baits were assigned a regulatory category (-1, 0, 1), based on the $\log_2$ fold-change in SLAM-seq read count between the control and SCC1-depleted samples. If the $\log_2$ fold-change was ≤ -0.1 and the FDR was ≤ 0.05, the regulatory category was -1. If the $\log_2$ fold-change was ≥ -0.1 and ≤ 0.1 and the FDR was > 0.05, the regulatory category was 0. If the $\log_2$ fold-change was ≥ 0.1 and the FDR was ≤ 0.05, the regulatory category was 1.

Ordinal logistic regression was then performed on the dependent and both independent variables using the polr() function from the MASS package (Venables *et al.*, 2002). The proportional odds assumption was tested by performing a graphical parallel slopes test.

2.13.5 Ordinal logistic regression: compound rewiring and compaction versus SLAM-seq signal

DESeq2 (Love *et al.*, 2014) was used to detect differential TN5 integration events per RF between conditions: SCC1+ and SCC1-. After DESeq2 analysis, RFs showing a baseMean value ≥ 1000 were discarded, removing 25 RFs (<0.02%). The DESeq2 results were used to categorise RFs into de-compacted, constant and compacted (-1, 0 and 1 respectively). If the $\log_2$ fold-change was > 0 and the FDR

was ≤ 0.05, the compaction category was -1. If the $\log_2$ fold-change was > -0.3 and < 0.3 and the FDR was > 0.75, the compaction category was 0. If the $\log_2$ fold-change was < 0 and the FDR was ≤ 0.05, the compaction category was 1. Additionally, a cut-off score was used to select accessible or inaccessible RFs (as opposed to a *change* in compaction). This score was set at the 10^{th} and 90^{th} percentiles of the mean ATAC-seq score per RF, providing cut-off values at -1.28 and 1.02 for compaction and decompaction respectively. The independent variable (X) was then calculated by performing the following steps in sequence on the interaction level:

1. Set the default value X = 0
2. If a promoter interaction was *lost* and the ATAC-seq score at the PIR was > 1.02 before cohesin depletion, set X = -1
3. If a promoter interaction was *maintained* and the ATAC-seq score at the PIR was > 1.02 after cohesin depletion, and the PIR is classified as constant or decompacted, set X = 1
4. If a promoter interaction was *maintained* and the ATAC-seq score at the PIR was > 1.02 before cohesin depletion, and the PIR is classified as decompacted, set X = 1
5. If a promoter interaction was *maintained* and the ATAC-seq score at the PIR was < -1.28 after cohesin depletion, and the PIR is classified as compacted, set X = -1
6. If a promoter interaction was *gained* and the ATAC-seq score at the PIR was > 1.02 after cohesin depletion, set X = 1
7. If a promoter interaction was *gained* and the ATAC-seq score at the PIR was < -1.28 after cohesin depletion, and the PIR is classified as compacted, set X = -1

This provides a value of X for each promoter interaction. The sum of the X values per bait is then taken. This value per bait is subsequently restricted to the range: [-1, 1]. I.e., if the sum of X-values for a given bait is 3, it is set to 1. If it is -3, it is set to -1. This results in a value of the independent variable X which can take on the values [-1, 0, 1].

The dependent variable was the SLAM-seq category per bait (see previous section). A treatment contrasts matrix was used with the dcv = 0 category as the baseline

group. Ordinal logistic regression was performed using the polr() function from the MASS package (Venables *et al.*, 2002).

2.14 Calculating ChIP-Seq score overrepresentation at baits and PIRs

This analysis included ChIP-Seq targets that were found with LASSO logistic regression to be predictors of *maintained* and *lost* promoter interactions. In total, 21 ChIP-Seq targets were selected. For each ChIP-Seq target separately, the 95th percentile score cut-off was calculated. This cut-off was used to categorise RFs as "strongly bound" or "not strongly bound" by a ChIP-Seq target. The same cut-off was used for baits and PIRs. The total number of RFs per rewiring category (*lost*, *maintained* or *gained*) with "strongly bound" and "not strongly bound" target was computed. This results in a 2x3 contingency table for each ChIP-Seq target (one for baits and one for PIRs). As an example, H3K4me3 shows the following tallies at PIRs:

Table 1. Example contingency table of H3K4me3 at PIRs.

	Lost	*Maintained*	*Gained*
Strongly bound	204	281	30
Not strongly bound	24759	8483	1582

Subsequently, Fisher exact tests were performed on these contingency tables, followed by Benjamini & Hochberg multiple testing correction across all tests. Additionally, log-odds ratios were calculated for baits and PIRs separately for each ChIP-Seq target, which were visualised as a heat map in Figure 50.

2.15 Bootstrap analysis on SLAM-seq data

Up- and downregulated genes upon cohesin depletion were selected using the 99th percentile of the absolute $\log_2$ fold-change in SLAM-seq read count between control and SCC1-depleted cells as a cut-off, corresponding to a value of 1.50 ($\log_2$ fold-change ≤ -1.50 for down-regulated genes and $\log_2$ fold-change of ≥ 1.50 for up-regulated genes). Additionally, the FDR was required to be ≤ 0.05. To select genes that were non-regulated upon cohesin depletion, the absolute $\log_2$ fold-change was required to < 0.1, the FDR was required to be > 0.05 and the gene was required to show an RNA-Seq RPM value ≥ 75th percentile in the control (non-cohesin-depleted) sample, corresponding to a value of 0.21.

Subsequently, contingency tables were constructed for *lost*, *maintained*, and *gained* promoter interactions separately. A pseudocount of 1 was added to avoid division by 0. As an example, *maintained* interactions show the following tallies of promoter interactions:

Table 2. Example contingency table of rewiring promoter interactions in the context of transcriptional regulation upon cohesin depletion.

	Up-regulated	Non-regulated	Down-regulated
Lost	22	1127	182
! *Lost*	41	680	44

Promoter interaction rewiring and transcriptional response were then expressed as a log odds ratio:

(5) $$\ln\left(\frac{(nLost \wedge downRegulated / n!Lost \wedge downRegulated)}{(nLost \wedge !downRegulated / n!Lost \wedge !downRegulated)}\right)$$

This ratio is > 0 if genes with lost promoter interactions tend to be more strongly down-regulated than genes with maintained or gained promoter interactions. Note that these log odds ratios were calculated for all combinations of rewiring category and regulatory category (e.g., lost & up-regulated, lost & non-regulated, etc). An empirical null distribution of log-odds ratios was then computed for each of these combinations, by constructing the contingency tables on randomly sampled baits (sampled from the set of baits that were categorised as described above) over 10,000 iterations. Subsequently, empirical p-values were calculated by dividing the number of times the values in the expected distribution exceeded the observed value by the total number of values in the expected distribution. The p-values were converted to z-scores by subtracting the expected log odds ratio from the observed log odds ratio and dividing this difference by the standard deviation of the distribution of expected log odds ratio. These z-scores were visualised as a heat map in Figure 53.

2.16 Calculating transcriptional difference between genes whose connections with the same enhancer are differentially rewired upon cohesin depletion

To select active RFs, ChIP-Seq scores per RF were used. RFs were marked active if the score for H3K4me1 was ≥95th percentile while simultaneously showing ≥95th percentile scores for H3K27ac *or* H3K4me3. Only active RFs were included in this analysis.

Subsequently, promoter interactions were selected where a PIR was contacted by multiple baits and the interactions were *lost* and *maintained* upon cohesin depletion, respectively. In the resulting set of promoter interactions, baits that interacted with multiple PIRs were discarded. Then, SLAM-seq $\log_2$ fold change values were obtained for these baits and baits lacking SLAM-seq data were discarded. Using the t.test() function, a one-sided t-test was performed on the $\log_2$ fold change scores (baits with *lost* versus baits with *maintained* promoter interactions) with the alternative hypothesis set to “less”.

2.17 Principal component analysis of ATAC-seq data

Principal component analysis (PCA) was performed in SeqMonk (https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Reads were stratified into 10kb bins with a 5kb step. Reads per bin were then log-transformed, corrected to the largest datastore and submitted to PCA.

2.18 Experimental analyses

Please note that all experimental analyses described in this thesis, except ATAC-seq, were performed by collaborators.

2.18.1 WAPL/PDS5A/PDS5B RNA interference

HeLa cells were transfected with siRNAs as described in (Lelij *et al.*, 2014) before adding thymidine. The siRNA sequences were obtained from Ambion and their compositions are: WAPL (50-CGGACUACCCUUAGCACAAtt-30), PDS5A (50-GCUCCAUAUACUUCCCAUGtt-30), PDS5B (50-GAGACGACUCUGAU CUUGUtt-30), NIPBL (50-GCAUCGGUAUCAAGUCCCAtt-30) and SMC2 (50-UGCUAUCACUGGCUUAAAUtt-30). Transfection was performed by incubating duplex siRNA with RNAi-MAX reagent (100 nM) in growth medium lacking antibiotics. Cells were harvested after 72 h of RNAi treatment.

2.18.2 Cell cycle synchronisation

HeLa Kyoto cells were synchronized at the G1/S-phase transition by two consecutive cell cycle arrest phases using 2 mM thymidine and released into fresh medium for 6 h (G2-phase), or 15 h (G1-phase). For mitotic cells, Nocodazole (100 ng/ml) was added 8 hours after release from double-thymidine block, to arrest the cells in prometaphase. Post-mitotic cells were removed by shake off after five hours.

2.18.3 Auxin induced degradation of SCC1 and CTCF

The experimental procedures were performed as described in (Wutz et al., 2017). Briefly, HeLa Kyoto cell culturing was performed as in (Nishiyama *et al.*, 2010). C-terminal tagging SCC1 and CTCF with an Auxin induced degron (AID) was performed using CRISPR/Cas9-mediated genome editing with a double-nicking approach (Ran *et al.*, 2013). The C-termini were extended with monomeric EGFP (L221K) and the IAA1771- 114 (AID*) mini-degron from *Arabidopsis thaliana* (Morawska *et al.*, 2013). Single clones were selected using flow cytometry PCR was used to confirm that all alleles were successfully modified.

2.18.4 Promoter capture Hi-C

The full description and visual demonstration of the protocol is given in (Schoenfelder *et al.*, 2018). Briefly, Hi-C library preparation was followed by a promoter capture step using SureSelect target enrichment according to the manufacturer's instructions (Agilent Technologies), using a biotinylated RNA bait library and custom paired-end blockers.

2.18.5 SLAM-seq

SLAM-seq was performed according to (Herzog *et al.*, 2017), following computational processing using SLAM-DUNK pipeline (https://t-neumann.github.io/slamdunk/)

2.18.6 ATAC-seq

ATAC-Seq (assay for transposase accessible chromatin) was performed according to the original protocol (Buenrostro *et al.*, 2013). ~71,000 cells were harvested, following centrifugation for 5 minutes at 3000 rpm. Cells were then re-suspended in 200µl ice cold lysis buffer for 15 minutes (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl2, 0.4% Igepal). The resulting nuclei were centrifuged for 10 minutes at 3000 rpm. The pellets were placed on ice and were then subjected to the transposase reaction: 25µl 2X TD buffer, 2.5µl Tn5 transposase (Nextera DNA library preparation

kit – Illumina FC-121-1030), and 22.5µl nuclease-free water. Nuclei were incubated for 30 min at 37°C and the DNA was subsequently purified using a Qiagen MinElute kit and then put into 10µl elution buffer, prior to library amplification.

Eluted DNA was added to a PCR reaction consisting of 5µl transposed DNA, 3µl Milli-Q, 1µl of 10µM mixed primers, 0.9µl 10x SYBR Green I and 5µl NEBNext High-Fidelity 2x PCR master mix. Reactions were amplified for 1 cycle at 98°C for 30 seconds, following 25 cycles at 98°C for 10 seconds, 55°C for 30 seconds and 72°C for 60 seconds. ATAC-seq libraries were subjected to bioanalyzer and KAPA qPCR analysis to test library concentration and fragment-size profiles. Sequencing was performed on an Illumina HiSeq 2500 in 50 bp paired-end sequencer in high output mode. Bioanalyzer analysis and sequencing was performed by the Babraham Next Generation Sequencing Facility.

After sequencing, FastQC was used to perform quality control on the DNA fragments. Subsequently, ATACseqQC (Ou *et al.*, 2018) was used to perform ATAC-seq-specific QC, confirming increased coverage at the TSS and an anti-correlation of nucleosome free and mononucleosome signals.

3 Investigating the interplay between promoter interactions and topologically associating domain organisation

To better understand the mechanisms that govern promoter-enhancer interactions and by extension their role in transcriptional control, it is paramount to expand our knowledge about the relationship between “focused” promoter interactions and global TAD organisation. Importantly, the localisation of promoter interactions with respect to TAD boundaries is poorly understood. Furthermore, it is unknown to what extent promoter interactions can cross TAD boundaries. In this chapter I present a joint genome-wide investigation of promoter interactions and TAD organisation. Additionally, I investigate transcriptional activity, TF binding and histone modifications at promoter-interacting loci in relation to TAD organisation.

3.1 Analysis of promoter interactions and TAD partitionings from 8 human haematopoietic tissues

TADs and promoter interactions describe separate, but related levels of spatial genome organisation. However, the precise influence of TADs on promoter interactions (and by extension transcriptional regulation) is poorly understood. A

prominent outstanding question is that of the insulating properties of TAD boundaries on promoter interactions.

To investigate this, I started by analysing a comprehensive collection of Hi-C and PCHiC datasets from 8 human haematopoietic tissues (median number of unique captured read pairs: 668,675,248; median number of detected promoter interactions: 186,172) (Javierre *et al.*, 2016). Firstly, I aimed to show what proportion of promoter interactions are not insulated by TAD boundaries and what this proportion would be if TAD organization would be random. To increase the robustness of TAD boundary detection, I constructed reference TAD partitionings by merging TAD calls across 2 replicates for each cell type, requiring a TAD to be detected in both replicates.

The number of detected TADs per sample ranged from 3724 in neutrophils to 5020 in macrophages, while the number of called TADs was similar between replicates (see Table 3).

I next quantified agreement between replicates by calculating the Jaccard index between pairs of replicate TAD intervals, ranging from 1 (complete overlap) through 0.5 (overlap spanning half the combined length of TADs) to 0 (no overlap). In all cell types, the majority of TADs showed overlap scores (expressed in the form of Jaccard indices) of at least 0.75. Figure 3 shows the distribution of Jaccard indices for Neutrophils and Macrophages. Based on the observed distributions, I opted for merging TADs with a Jaccard index of 0.75 or greater, defining the consensus boundaries as the midpoint of the boundary locations detected in each replicate (see Figure 4). This replicate merging approach produced high-confidence consensus TAD sets, which included a median of 2959 TADs (ranging from 2310 TADs in neutrophils to 3741 TADs in macrophages, Table 3). Mean TAD interval lengths ranged from ~450 kilobase pairs in Megakaryocytes to ~548 kilobase pairs in Neutrophils. Median TAD interval lengths ranged from ~363 kilobase pairs in Macrophages to 475 kilobase pairs in Neutrophils. These high-confidence TADs were then integrated with PCHi-C interactions as described below.

Table 3. Number of TADs called per replicate and results from TAD merging. The number of TADs per replicate is shown in columns 2 and 3. Column 4 shows the number of TADs with a Jaccard index ≥ 0.75

Tissue	TADs R1	TADs R2	Jaccard ≥ 0.75
Erythroid	4631	4570	3216
Macrophages	5020	4724	3741
Megakaryocytes	4964	4722	3541
Monocytes	3947	3938	2851
Naive B-cells	3877	3851	2940
Naive CD4 T-cells	3926	3779	2800
Naive CD8 T-cells	3839	3838	2978
Neutrophils	3778	3724	2310

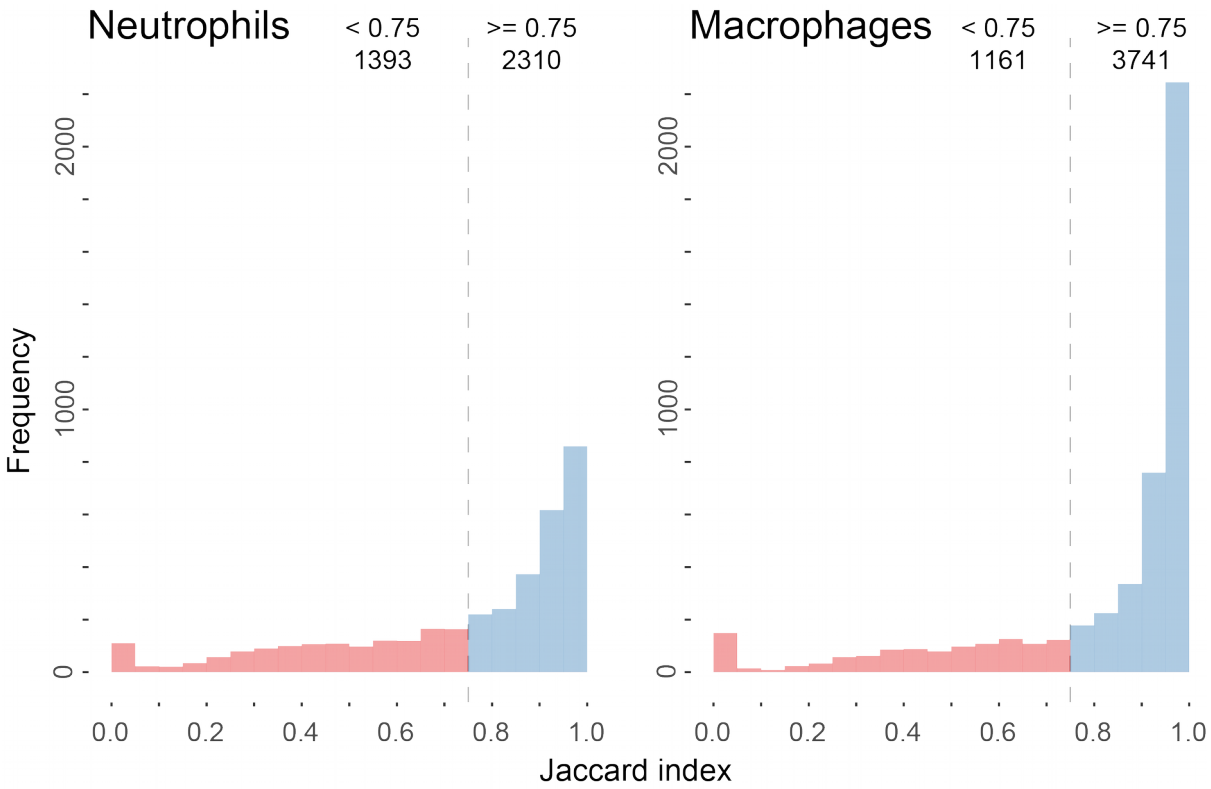

Figure 3. Merging replicate TAD partitionings discards TADs with discordant coordinates. Distribution of Jaccard index between paired TAD intervals from two replicates in Neutrophils (left) and Macrophages (right). TAD intervals with a Jaccard index >= 0.75 are shown in blue, those with a Jaccard index < 0.75 are shown in red. This figure shows the extremes among analysed tissues in terms of the proportion of TADs that meet the Jaccard index requirement: 61.6% for Neutrophils and 76.8% for Macrophages.

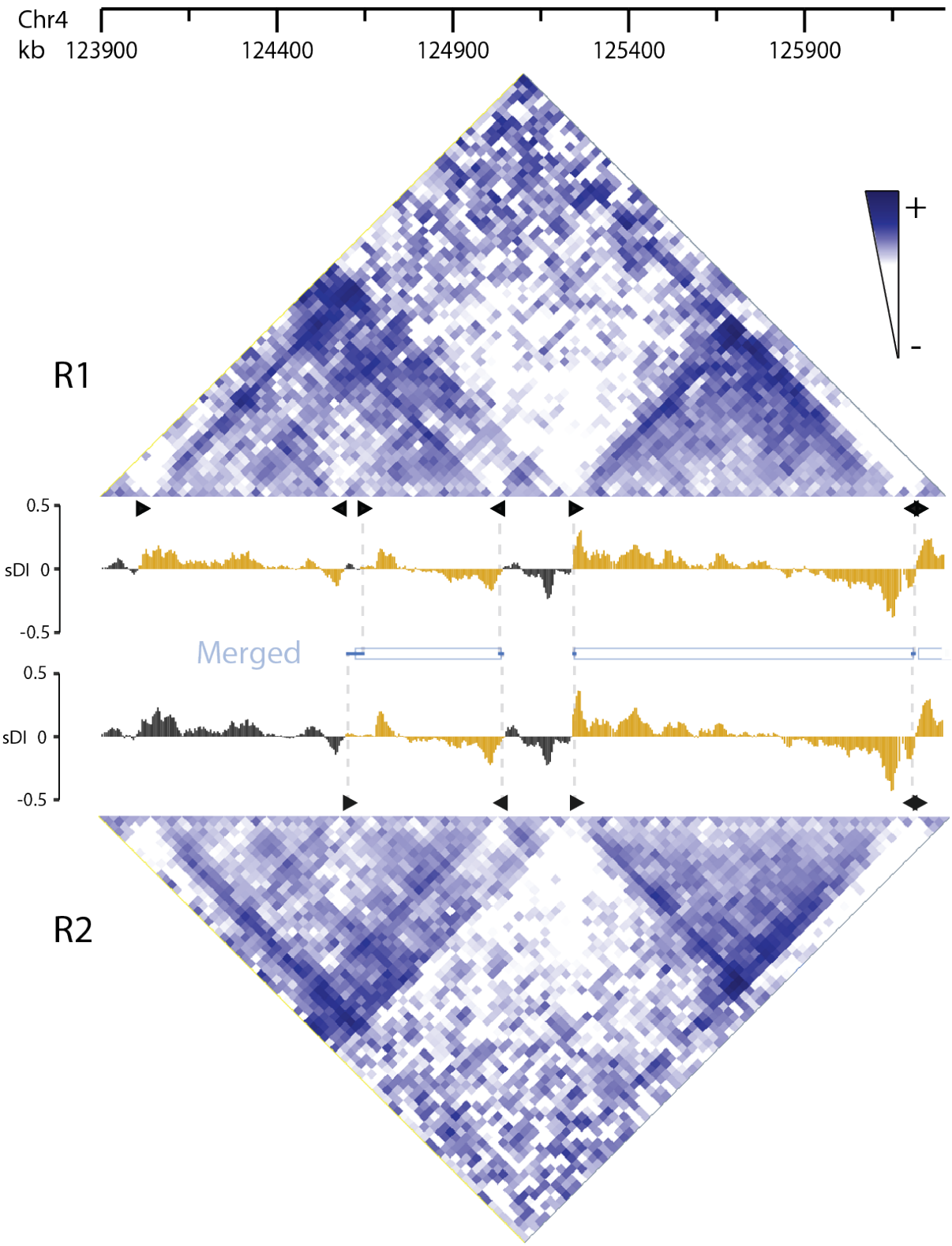

Figure 4. Representative illustration of comparing and merging TAD intervals from two replicates. This example shows data from biological replicates on chromosome 4 at coordinate interval [123,900,000 - 126,300,000] from HeLa cells. Hi-C contact matrices are binned at 40kb with a 30kb sliding window increment. The Hi-C contact matrices and

standardized directionality index (sDI) histograms show considerable similarity. In the first replicate (top) HOMER calls 3 TADs (yellow DI, demarcated by black triangles), whereas it calls 2 in the second replicate (bottom). Note that part of an additional TAD is visible on the far right of the figure. The consensus TAD partitioning is represented by blue rectangles. The hierarchical organisation of TADs is clearly visible in the contact matrices, where the called TADs appear as triangular shapes. The failure of HOMER to call the left most TAD in replicate 2 (R2) is potentially a false negative (which propagates into the merged set). The left-most consensus TAD boundary illustrates how the centre coordinate of the replicate TAD boundary coordinates is taken.

Having constructed sets of high-confidence TADs, I proceeded to investigate how strongly promoter interactions are insulated by their boundaries. To this end I quantified the proportion of promoter interactions that cross at least one TAD-boundary, in addition to the proportion of interactions per bait that cross at least one TAD boundary. I used interaction data that was pre-calculated with the Capture Hi-C analysis pipeline CHiCAGO (see section 1.3.4). I used a CHiCAGO score cut-off of 5 to define significant promoter interactions, as recommended (Cairns *et al.*, 2015).

I first asked what proportion of promoter interactions cross TAD boundaries, and compared this to random expectation. To compute the randomly expected proportion, I devised a strategy to compare the detected promoter interactions to a control set of randomised TADs (while preserving the population distribution of baits within TAD intervals) (see Methods 2.7).

The proportions of TAD-boundary crossing interactions ranged from ~27% to ~38% in Neutrophils and CD8+ T-cells, respectively (Figure 5). This was appreciably lower than the expected proportions in all eight tissues, which ranged from ~55% to ~67% in Neutrophils and Macrophages respectively. This result (published in Javierre *et al.* 2016) indicates that a substantial proportion of promoter interactions are not insulated by the detected TAD boundaries. Furthermore, the fact that the proportion of TAD-boundary crossing interactions is similar between several human cell types indicates that this result is likely general and not tissue-specific.

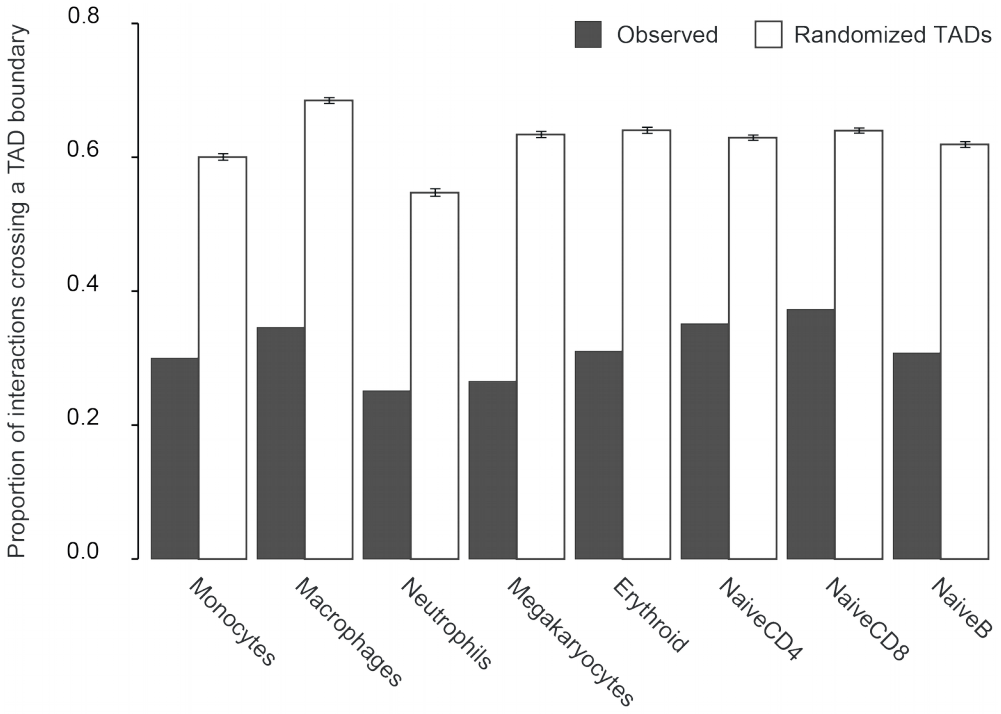

Figure 5. TAD boundaries insulate promoter interactions although a substantial proportion of interactions appear to cross at least one TAD boundary. Proportions of PCHiC interactions that cross ≥ 1 TAD boundary in 8 human haematopoietic tissues and upon TAD randomisation. Upon randomisation of TAD organisation (1000 permutations), the expected proportion of TAD boundary crossing promoter interactions is much higher. This indicates that promoter interactions are constrained when TAD organisation is present, though they may be insulated incompletely.

Since baits frequently interact with multiple PIRs, TAD boundary crossing on the interaction level (i.e. of pairwise interactions) provides an incomplete result. To investigate this, I asked the question: of all the interactions per bait, what proportion crosses TAD boundaries? I found that a considerable proportion of baits interacted exclusively within their respective TADs (Figure 6 and Table 4). However, baits that interacted both within and across their cognate TADs (mixed inter-TAD/intra-TAD) appear roughly as frequently as baits with intra-TAD-only interactions (See Table 4). Furthermore, an unexpectedly large minority of baits had all interactions crossing TAD boundaries (See Figure 6). Perhaps these promoter interactions reflect

hierarchical TAD organisation or perhaps they are supported by TAD-independent mechanisms.

I reasoned that if detection of inter-TAD-only promoter interactions is driven by TAD boundary-proximal baits, small misestimates of the exact location of the TAD boundary could produce this result as a technical artefact. To test this possibility, I repeated the same analysis, but this time restricted it to baits showing a central position within the TADs (40-60% of the TAD length). As can be seen from Figure 7, the resulting distribution is comparable to that observed in the non-restricted analysis (Figure 6). I therefore conclude that the detected “inter-TAD-only” baits were unlikely to arise as a result of this technical error.

Taken together, these results indicate that promoter interactions are constrained by, but not fully contained within TAD boundaries. I further investigate the baits with inter-TAD-only interactions in sections 3.3 – 3.5.

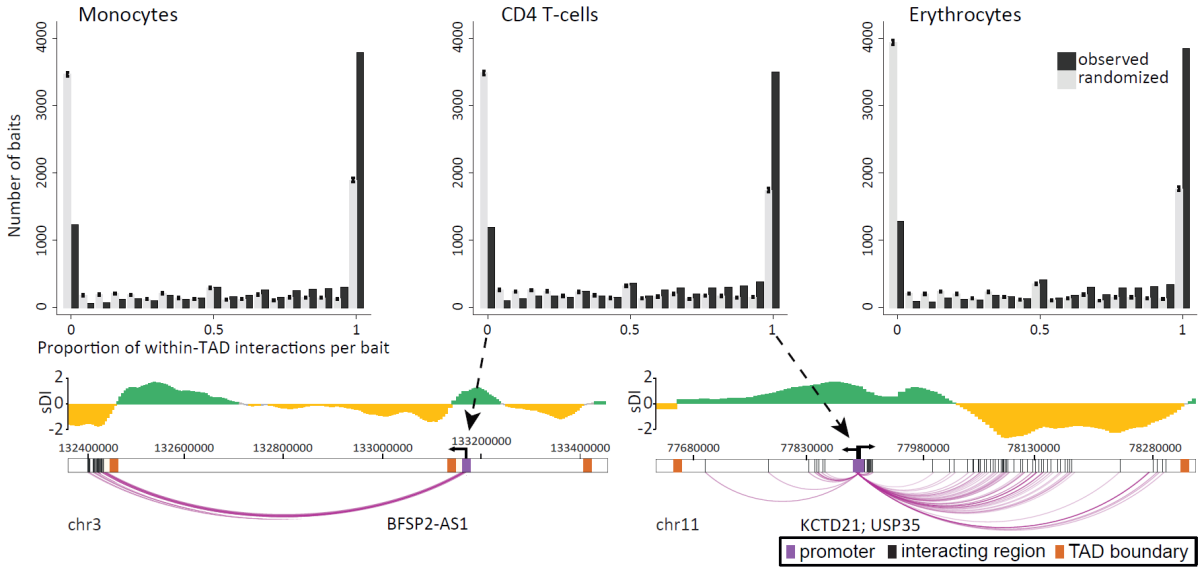

Figure 6. Distributions of the proportion of intra-TAD interactions per promoter. Top: proportion of within-TAD (intra-TAD) interactions per bait in three representative haematopoietic samples: Monocytes, CD4 T-cells, and Erythrocytes. Bottom: examples of directionality index, TAD boundary locations and promoter interactions. Baits predominantly interact intra-TAD but there are considerable numbers of baits that interact partially inter-TAD. A surprisingly large minority of baits interact solely inter-TAD. Randomised TADs show many fewer baits with intra-TAD-only interactions and many more baits with inter-TAD-only interactions. Error bars represent mean +/- 1SD over 1000 permutations.

Table 4. Number of baits by boundary crossing category: intra-TAD-only, mixed inter-TAD/intra-TAD, and inter-TAD only.

'Proportion in TAD' represents the proportion of interactions per bait that are located within the same TAD as the bait. E.g. when 'Proportion in TAD' equals 1, the bait has intra-TAD-only interactions.

	Proportion in TAD = 1	Proportion in TAD > 0 & < 1	Proportion in TAD = 0
Monocytes	3806	3327	1248
Macrophages	3806	4106	1583
Neutrophils	3704	2609	1040
Megakaryocytes	4455	3569	966
Erythrocytes	3862	3729	1297
CD4 T-cells	3514	4085	1207
CD8 T-cells	3682	4634	1502
B-cells	3952	4022	1346

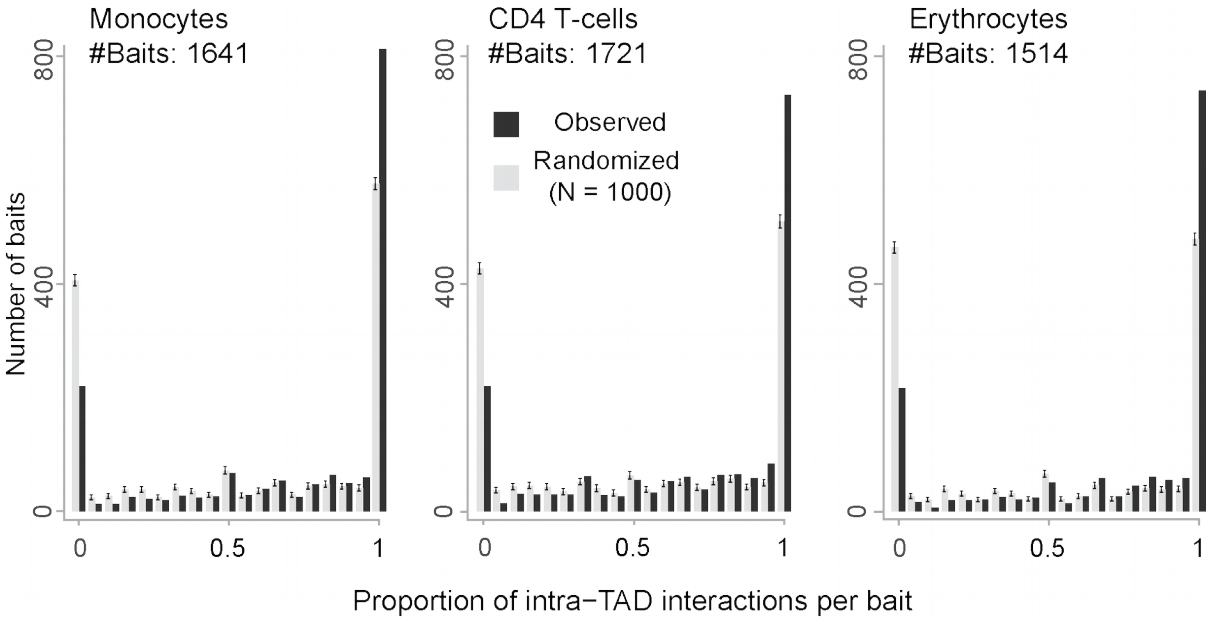

Figure 7. TAD-central baits show frequency profiles of inter/intra-TAD interactions similar to non-TAD-central baits. Distribution of the proportion of intra-TAD interactions per bait for three representative haematopoietic samples: Monocytes, CD4 T-cells and Erythrocytes. The bait position is restricted to the centre pentile (40-60%) of the TAD. The distribution of the proportion of intra-TAD interactions per bait is comparable to that of non-position-restricted baits. This shows that TAD-boundary crossing behaviour is similar for baits at the centre of a TAD compared to baits at any position within a TAD. Notably, baits with inter-TAD-only interactions are present at similar proportions to the non-position-restricted baits. Randomised TADs show fewer baits with intra-TAD-only interactions and many more baits with inter-TAD-only interactions. Error bars represent mean +/- 1SD over 1000 permutations.

3.2 Eliminating potential confounding through population heterogeneity by analysis of cell-cycle synchronized HeLa cells

Since the haematopoietic tissue samples are from nonsynchronous populations of cells, data obtained from these samples will reflect signals from the various stages of the cell cycle.

To ensure that cell cycle stage heterogeneity did not confound the TAD boundary calling procedure and the PCHi-C interaction calling procedure, I turned to a different model system: cell cycle-synchronised HeLa cells. Dr Gordana Wutz (Jan Michael Peters lab, IMP, Vienna) performed Hi-C and PCHi-C on HeLa populations synchronized in mitosis, G1 and G2, and Dr Csilla Varnai performed initial Hi-C data processing, I performed additional Hi-C data processing, all PCHi-C data processing and QC, and the downstream analyses described below.

To define TADs and to produce Hi-C contact maps, we used the tool from the HOMER suite (Heinz *et al.*, 2010). I visually inspected the Hi-C contact maps, and I find that the interphase samples show great similarity (Figure 8). The mitotic data however, show low signal intensity and no clear TAD organisation, consistent with previous reports (Naumova *et al.*, 2013). We produced TAD calls with two stringency settings (see Methods 2.2) referred to hereinafter ‘standard’ and ‘stringent’ TADs respectively. In G1, we detected 4456 standard and 2585 stringent TADs. In G2, we detected 4641 standard and 2668 stringent TADs. In mitotic cells the decrease was most pronounced with 1200 standard and 532 stringent TADs. Table 5 provides an overview of the numbers of TADs per sample. Mean standard TAD interval lengths were ~405 kilobase pairs in G1 and ~395 kilobase pairs in G2. Median standard TAD interval lengths were ~328 kilobase pairs in G1 and ~333 kilobase pairs in G2. Mean stringent TAD interval lengths were ~570 kilobase pairs in G1 and ~552 kilobase pairs in G2. Median stringent TAD interval lengths were ~453 kilobase pairs in G1 and ~443 kilobase pairs in G2.

Notably, TAD intervals from mitotic cells showed low reproducibility: after merging TAD intervals (using the approach detailed above), the mitotic data only retained 27.6% to 35% of all called TADs (for normal and stringent settings respectively). This was in contrast to the more than 80% TAD intervals retained after merging in both G1 and G2 for both standard and stringent TADs (Figure 9).

I then calculated genome-wide Jaccard index values between all combinations of paired samples (Figure 10). Note that this is a global Jaccard index ($bp_{A,B}/bp_{A,B}$ where A and B are samples), different from the TAD-wise Jaccard indices presented previously ($bp_{Ta,Tb}/bp_{Ta,Tb}$ where Ta and Tb are TADs). Figure 10 shows the genome-wide jaccard index between pairs of samples and for two stringency settings. This further illustrates that TAD partitionings derived from cells in Mitosis show little agreement with those in G1 and G2 and to a lesser extent also between replicates. The interphase samples on the other hand, show considerable agreement: in the high-stringency data at least 79.3% of all the partitioned genome intersects between G1 and G2 whereas this is 87.6% between G1 replicates and 87.8% between G2 replicates.

Since signal intensity and reproducibility are very low in the mitotic samples, I did not consider TAD organisation in those samples in the following sections. In contrast, since the interphase replicates showed considerable similarity, from section 3.4 onward (unless explicitly stated otherwise), I treated the four datasets for G1 and G2 jointly, generating reference TAD calls based on the two biological replicates for each of these cell cycle phases.

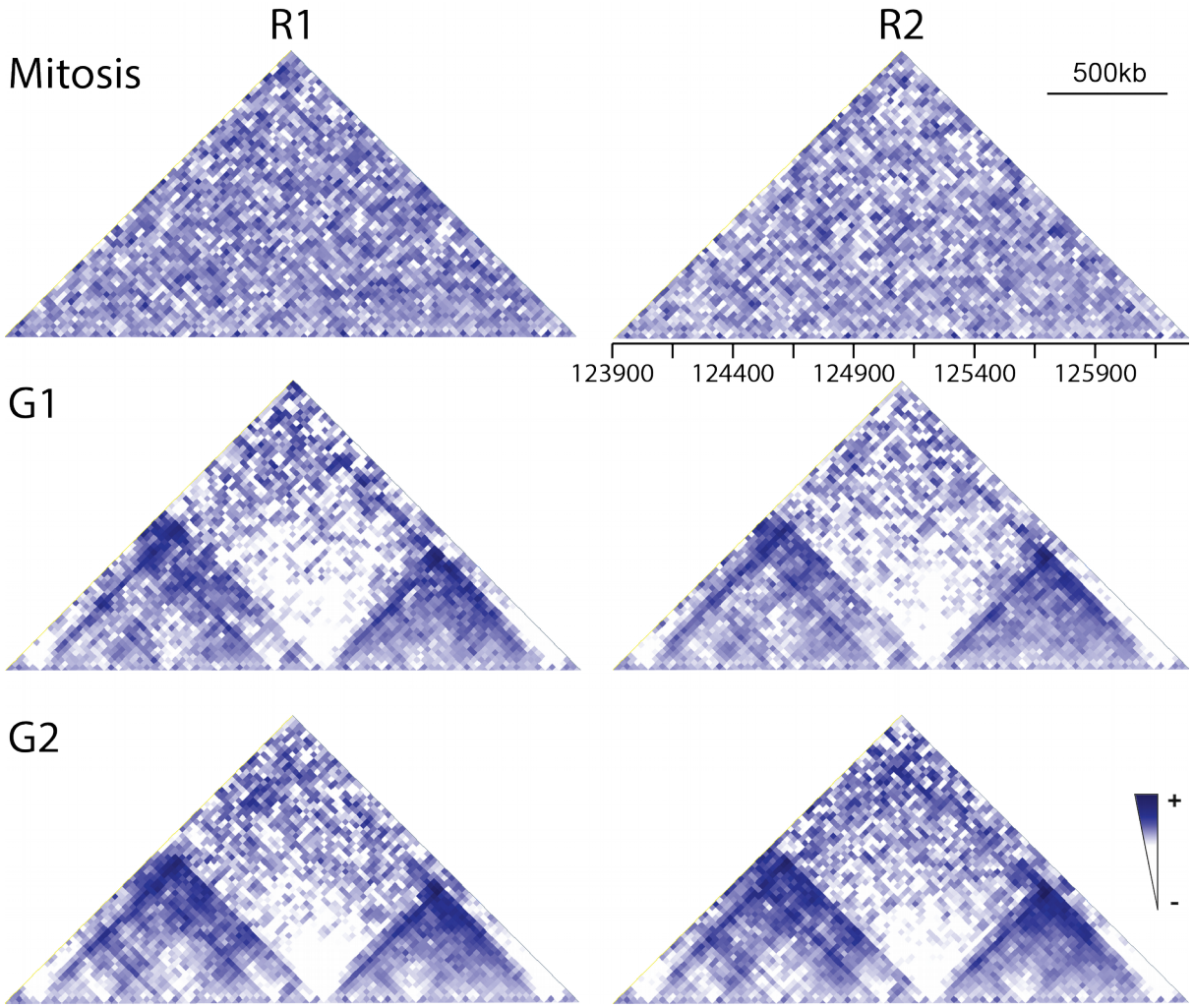

Figure 8. TAD organisation is absent in mitosis while it is preserved between G1 and G2. Representative example Hi-C contact matrices from synchronised HeLa cells in G1, G2 and Mitosis. Cells in G1 and G2 clearly show TAD structures whereas this is completely absent in Mitotic cells. The displayed region is on chromosome 4 [123900000, 126300000], Hi-C contact matrices are of 40kb bins with a 30kb sliding window increment.

Table 5. Number of TADs called per replicate and results of TAD merging. The number of TADs per replicate is shown in columns 2 and 3. Column 4 shows the number of TADs with a Jaccard index ≥ 0.75

Tissue	TADs R1	TADs R2	Jaccard ≥ 0.75
Mitosis standard	4407	4281	1200
Mitosis stringent	1543	1495	532
G1 standard	5378	5328	4456
G1 stringent	3028	2985	2585
G2 standard	5530	5515	4641
G2 stringent	3115	3104	2668

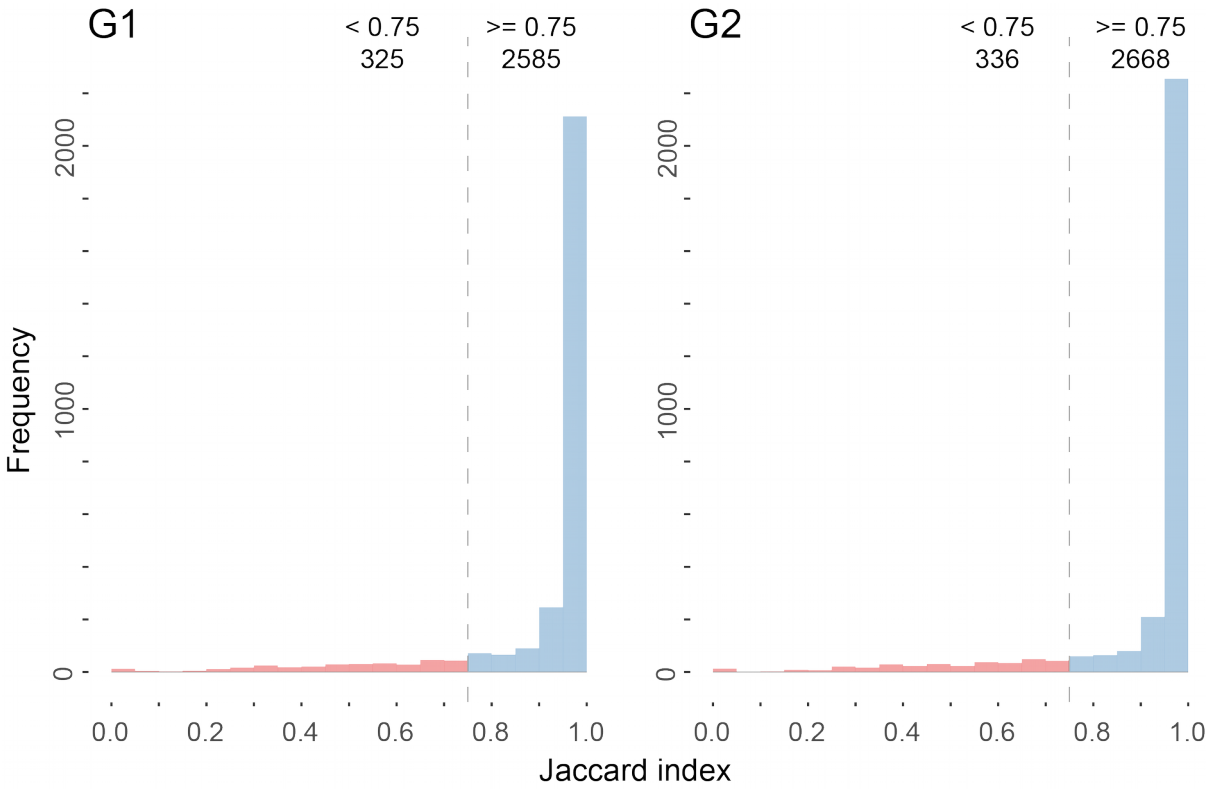

Figure 9. Synchronised HeLa cells show greater consensus between replicate TAD partitionings than the hematopoietic tissues. Distribution of jaccard index between stringent TAD intervals from two replicates in G1 (left) and G2 (right). TAD intervals with a jaccard index ≥ 0.75 are shown in blue, those with a jaccard index < 0.75 are shown in red.

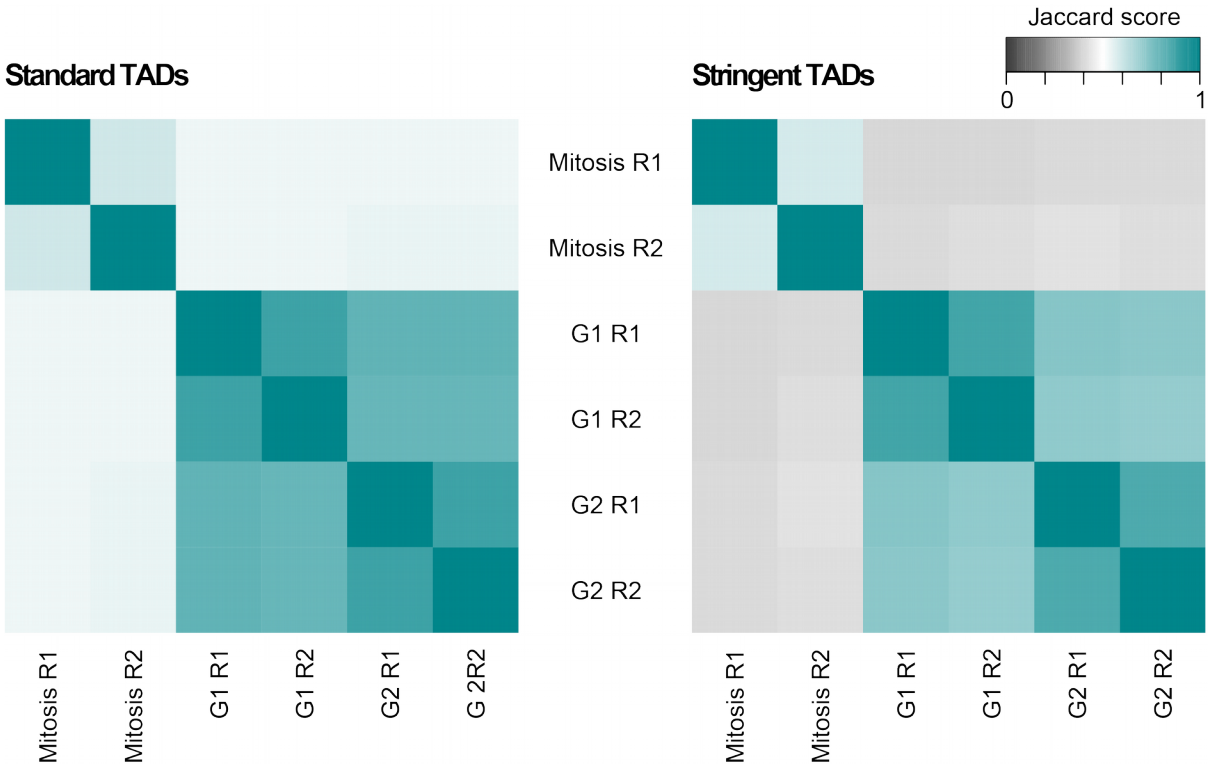

Figure 10. TAD partitionings show greater similarity in interphase than in Mitosis. Jaccard indices over genome-wide TAD partitionings are shown for bothreplicates and for two TAD stringency settings. Interphase partitionings show considerable similarity, with jaccard indices of 0.876 between G1 replicates and 0.878 between G2 replicates, and at least 0.793 between cell cycle stages in G1R2 vs G2R2. The interphase partitionings show little similarity to the samples in Mitosis. Additionally, the Mitotic samples show low similarity between replicates. This indicates that the partitionings in the Mitotic samples are irreproducible and that called TADs in these samples are likely false positives.

3.3 Analysis of promoter interactions and TAD partitionings from synchronised HeLa cells

Having established that TAD organisation is broadly invariant between G1 and G2, but is weakened in mitosis, I turned to investigating the cell cycle dynamics of promoter interactions detected with PCHi-C in the same three cell cycle-resolved populations of HeLa cells. Using CHiCAGO, I detected 147402 and 144434 significant promoter interactions (score ≥ 5) in G1 and G2, respectively. In mitosis I detected significantly fewer interactions: 30111, which reflects the situation in Hi-C presented in the previous section.

I asked whether promoter interactions detected in each of the analysed cell cycle phases showed distinct properties. To this end, I first compared the linear distance spanned by these interactions. Notably, promoter interactions detected in mitosis were considerably shorter-range than those observed in interphase cells, with a median distance of ~20kb compared with ~217kb in G1 and ~243kb in G2 (Figure 11).

I next compared the enrichment of promoter-interacting fragments (PIRs) detected in each cell cycle phase for chromatin marks (H3K4Me1, H3K27Ac, H3K27Me3 and H3K9Me3) and architectural proteins (cohesin subunit SMC3 and CTCF), using publicly available ChIP-seq data from ENCODE (Encode Consortium, 2013). As can be seen in Figure 12, while PIRs in G1 and G2 showed enrichment for both SMC3 and CTCF, this enrichment was considerably weaker at PIRs detected in mitosis, consistent with mitotic chromatin organisation and the support thereof by condensin rather than cohesin (Kalitsis *et al.*, 2017). In addition, interphase PIRs, but not mitotic PIRs were enriched for the enhancer-associated mark H3K4Me1, and the active chromatin mark H3K27Ac. In contrast, the repressive marks H3K27Me3 and H3K9Me3 did not appreciably exceed what is expected random at the PIRs in all three samples.

To further investigate the dynamics of promoter interactions between the cell cycle stages, I performed K-means clustering on CHiCAGO interaction scores (see Methods for details). The resulting partitioned data are visualised as a heat map in Figure 13. This shows clearly that the largest cluster (A) corresponds to interphase promoter interactions. This cluster accounts for 45.4% of all significant interactions detected across all included cell cycle stages. In contrast, cluster B demarcates cell

cycle-invariant interactions, which account for just under a quarter of all promoter interactions (24.5%). The following two clusters (C - 12.6% total; D - 10.1% total) contain PCHiC interactions with high scores in either G1 or G2, respectively. Notably, promoter interactions in these clusters still show appreciable scores in the "other" interphase stage (i.e. cluster C interactions in G2 and cluster D interactions in G1, respectively). Finally, cluster E shows mitosis-specific promoter interactions, accounting for 7.3% of the total number. Taken together, these results indicate that while the majority of promoter interactions interphase-specific and absent in mitosis, a significant proportion of promoter interactions are detected throughout the cell cycle. Furthermore, these results identify small sets of promoter interactions that are specific to a single cell cycle stage, including mitosis.

I next sought to investigate the relationship between TAD boundaries and promoter interactions in a cell-cycle-specific fashion, following the methodology established in section 2.7 (I restricted this analysis to G1 and G2 only, given the low signal strength and low reproducibility of TADs in mitosis). I paired PCHiC interactions from G1 with consensus TAD-intervals from G1 and interactions from G2 with consensus TAD-intervals from G2, performing these analyses with two TAD-calling stringency settings (see Figure 14).

On the interaction level, the proportion of promoter interactions in G1 that cross TAD boundaries (~29%) was slightly lower than in G2 (~34%) respectively (standard TADs). This difference remained with the more stringent TAD calls, showing ~22% and ~25% TAD-crossing interactions in G1 and G2, respectively. These results may point to a slightly higher insulating power of G1 compared with G2 TADs.

On the per-bait level, the distribution of the proportion of standard TAD-boundary crossing interactions is comparable to the haematopoietic samples (compare Figure 6 and Figure 15). However, using stringent TADs, the number of baits with inter-TAD-only interactions is roughly halved. This highlights a distinction between the interaction level and the per-bait level, in terms of the response to increased stringency of TAD detection: a reduction of ~8% on the interaction level compared to a reduction of ~50% on the per-bait level.

An additional implication of the above results is that the more stringently called TADs are less penetrable to promoter interactions, suggesting that some of the weaker

TADs potentially correspond to a lower level of chromosomal domain hierarchy, such as sub-TADs. Additionally, this may indicate that some of the standard TAD calls are false-positives. For these reasons, I focused on the stringent TAD calls in further analyses.

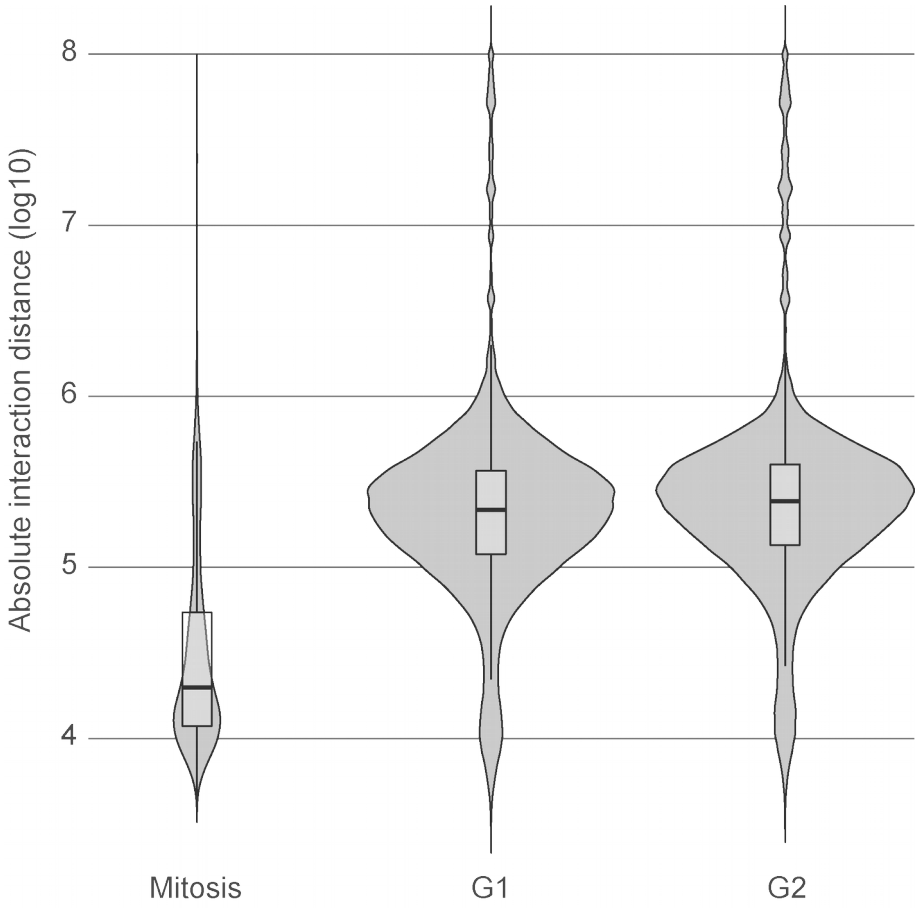

Figure 11. Promoter interaction distance distribution in G1 is similar to G2 and both show longer-range interactions than in mitosis. Violin plots showing the PCHi-C interaction distance in samples: mitosis, G1 and G2. PCHi-C interactions from mitotic cells clearly show a shorter interaction distance. In addition, the mitotic sample shows far fewer interactions than the interphase ones. PCHi-C interactions in G1 and G2 are similar in terms of interaction distance and number. The surface area of the violins reflects the number of interactions: 21,243 (mitosis), 119,913 (G1), 116,890 (G2). Only interactions with a CHiCAGO score of ≥ 5 were considered. Bait-to-bait interactions and trans-chromosomal interactions were discarded.

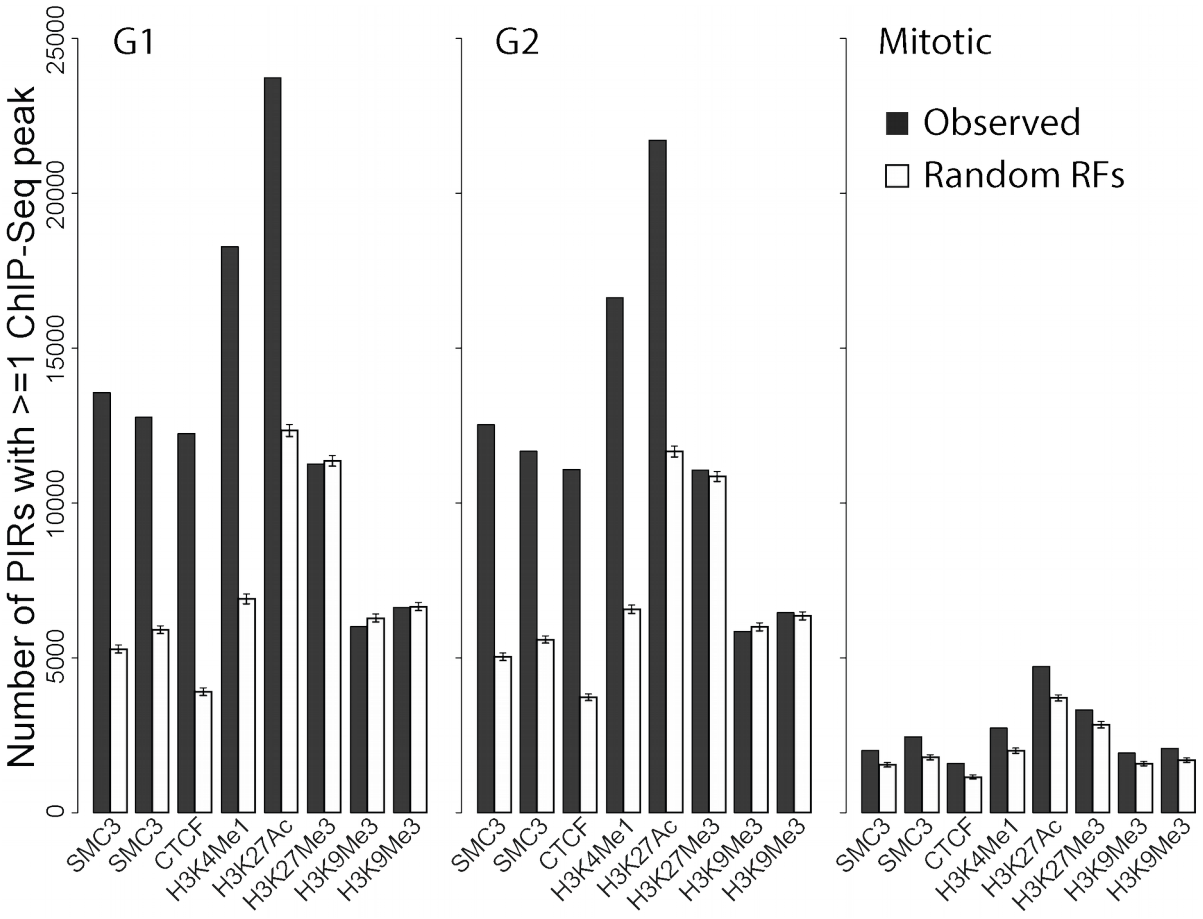

Figure 12. Promoter interactions in G1 and G2 show similar ChIP-Seq peak frequency profiles but those in mitosis are dissimilar. CHiCAGO ChIP-Seq frequency plot. Frequency of significant ChIP-Seq peaks at PIRs and at randomly selected restriction fragments in G1, G2 and Mitosis. Interphase PIRs show clear enrichment for Cohesin (SMC3), CTCF, as well as active enhancer marks H3K27Ac and H3K4Me1. No enrichment is found for heterochromatin marks H3K27Me3 and H3K9Me3. PIRs from Mitotic cells show mild enrichment for all targets. It should be noted that the ChIP-Seq data derive from public datasets on unsynchronised HeLa cells.

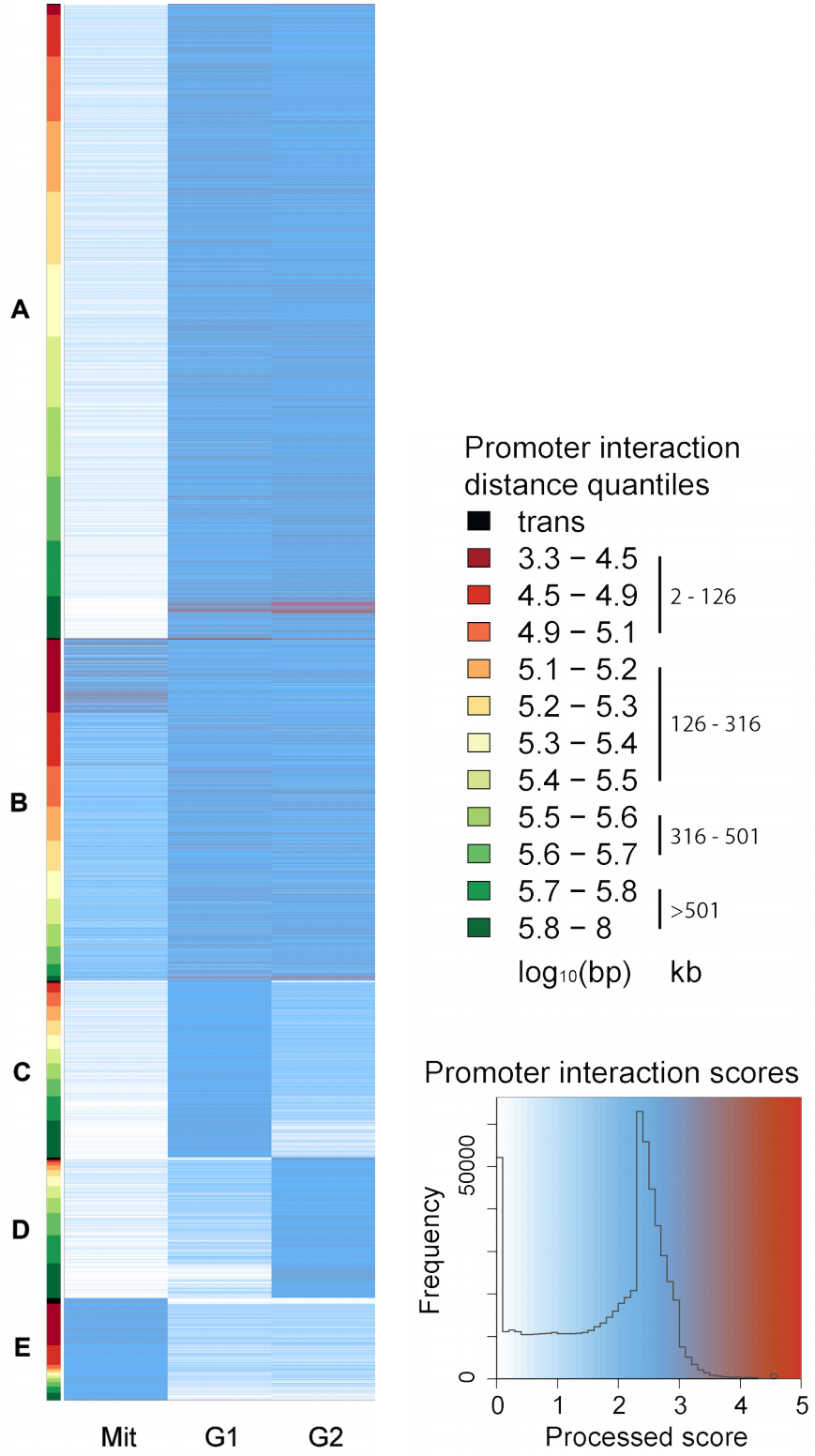

Figure 13. Clustered processed CHiCAGO scores from cells in Mitosis, G1 and G2. Cluster A shows interactions in G1 and G2 and contains 94293 promoter interactions, which represents 45.4% of all interactions. Cluster B shows interactions in all three samples, it has 50879 interactions which represents 24.5% of all interactions. Cluster C

shows interactions in G1 and weak interactions in G2, it has 26219 interactions which represents 12.6% of all interactions. Cluster D shows interactions in G2 and weak interactions in G1, it has 20864 interactions which represents 10.1% of all interactions. Cluster E shows interactions in Mitosis and weak interactions in G1 and G2, it has 15234 interactions which represents 7.3% of all interactions. Promoter interaction distance has been colour coded (left) and ranges from short-range (red; 2-126kb) to long-range (green; >501kb).

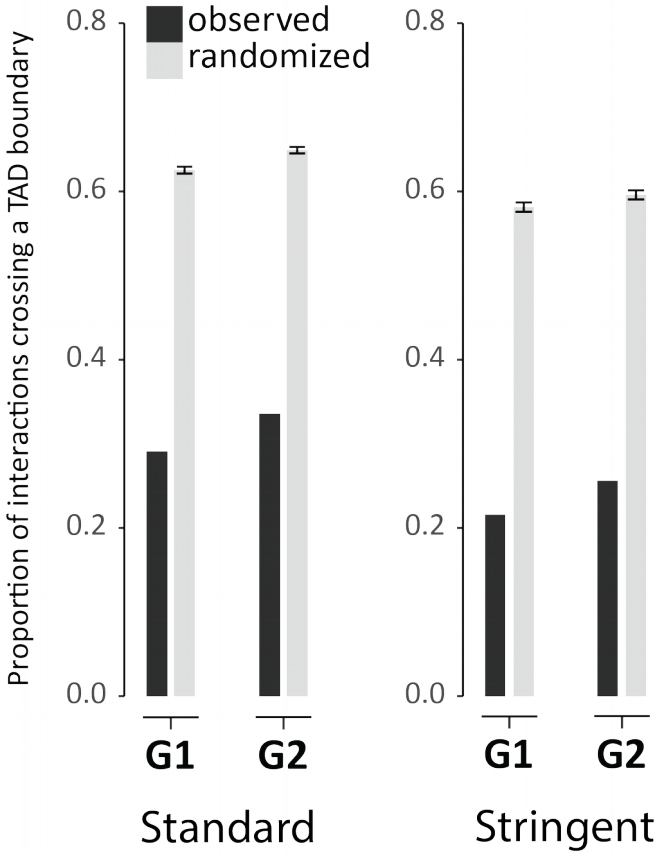

Figure 14. The proportion of promoter interactions that cross ≥ 1 TAD boundary is similar between G1 and G2. 29.2% cross in G1 and 34.0% cross in G2 using ‘standard’ TADs, and 22.2% in G1 and 25.4% in G2 using ‘stringent’ TADs. In both cases, the TAD boundary crossing proportion for randomized TADs is much higher. These boundary crossing properties are comparable to those measured in the haematopoietic tissues.

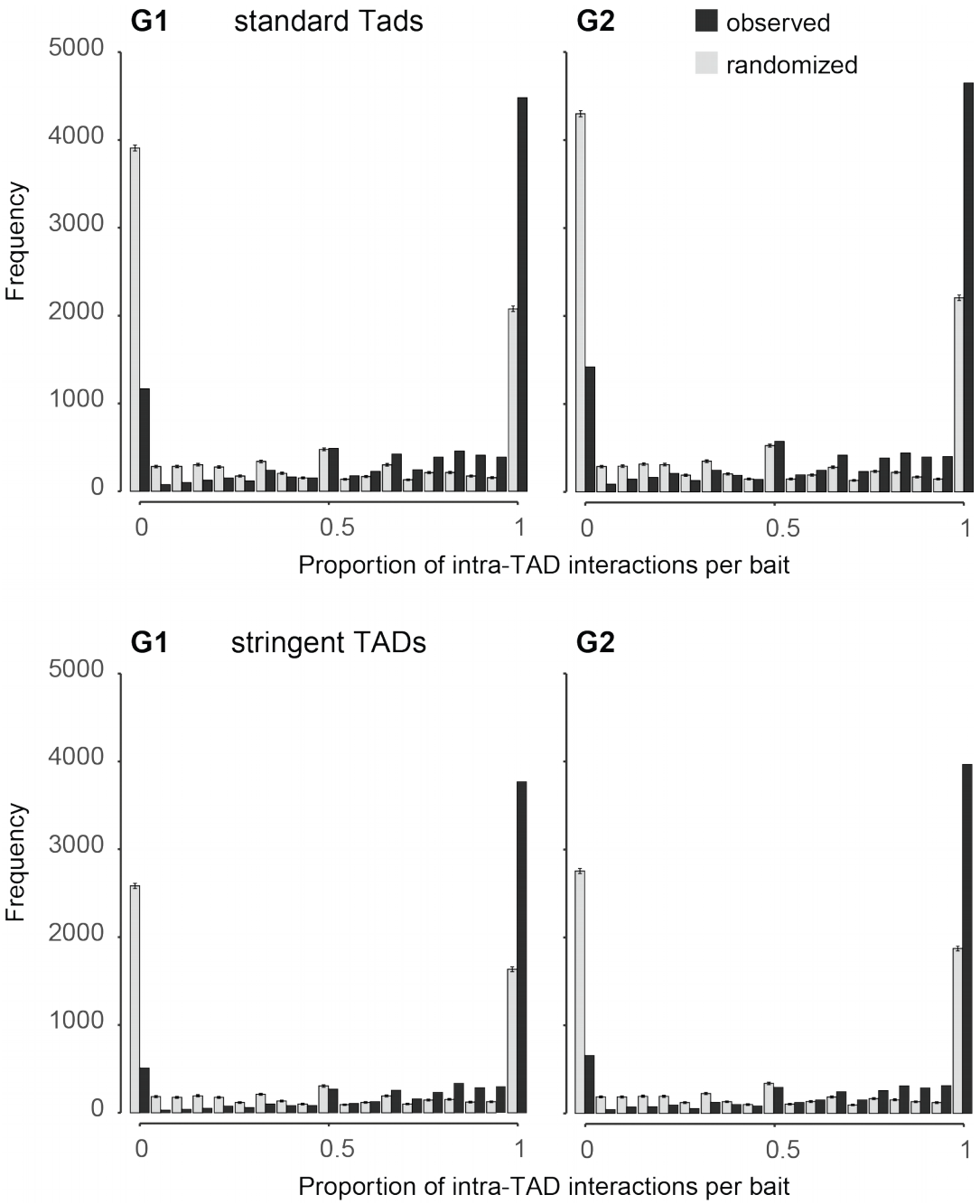

Figure 15. Distributions of the proportion of intra-TAD interactions per promoter in G1 and G2. Top: TADs called with ‘standard’ TADs, bottom: TADs called with ‘stringent’ TADs. The increased stringency setting results in a marked decrease in the detected number of “inter-TAD-only” baits.

3.4 Localisation of promoter interactions relative to TAD boundaries

To further investigate TAD boundary crossing of promoter interactions, it is paramount to detect the crossing events with high confidence. To this end, I devised a strategy that makes use of all four replicates of interphase TAD partitionings without merging them and is scalable to an arbitrary number of TAD partitionings. Note also that this approach is not limited to interactions of baits that are located within TAD intervals only. In this new approach, I integrated TAD intervals directly with RF intervals, by assigning the most proximal replicate TAD start and end coordinates to each RF. This enabled me to calculate the relative location of any RF with respect to its most proximal TAD and to calculate the variance over the four TAD partitionings.

Figure 16 shows the distribution of the relative locations of baits and PIRs within TADs for significant promoter interactions in G1 (left panel) and G2 (right panel), respectively. As can be seen from this figure, the frequency of baits and PIRs is relatively low beyond TAD boundaries. This shows that the majority of promoters and their PIRs are located in regions where TADs are defined. The figure also shows that within TADs, there is a strong tendency for baits and PIRs to be located proximally to TAD boundaries. In other words, the RFs that I detect to be involved in promoter interactions are not uniformly distributed within TADs but rather, are most commonly detected at TAD boundaries, supporting the notion that TADs are loop-like structures (Rao *et al.*, 2014). Lastly, the shaded area reveals little variance between the frequency distributions of the four interphase TAD partitionings, reflecting the previously observed similarities in interphase TAD organisation.

To further investigate the looping nature of promoter interactions in TADs, I asked the question: if a bait is located within a defined sub-region of a TAD, where are its PIRs located relative to the same TAD? To this end, I modified the previous analysis by limiting the relative position of one of the partner RFs to a pentile region (1/5th) of the TAD interval. For example, to analyse TAD boundary peripheral baits, I limit the relative position of the bait to the peripheral pentile of the TAD (coordinates [0, 0.2] and [0.8, 1]) and then plot the frequency of the PIR position (normalised to the bait-TAD). Figure 17 shows the resulting diagrams for three pentile “viewpoint windows”: peripheral, intermediate and central (note that the peripheral and intermediate windows combine data from the left as well as the right side of the TAD). This

reveals that peripheral RFs have a strong tendency to interact with a partner RF that is proximal to the opposite TAD-boundary, supporting the conceptualization of TAD as loops. The intermediate viewpoint shows a similar trend, although the amplitude of the bait frequency (shown in blue) is noticeably lower. The central window shows roughly symmetrical frequencies for both baits and PIRs. Taken together this analysis displays a preference of RFs to interact with a partner RF that is located proximally to the opposite TAD boundary, regardless of the position of the restricted RF within the TAD (boundary proximal, intermediate, central). For the peripheral window, this further supports the model of TADs as loops. However, the property of RFs located within TADs and away from the boundaries (i.e., within intermediate and central windows) to preferentially interact with partner RFs at the boundaries indicates that the looping nature of promoter interactions within TADs extends beyond boundary to boundary interactions.

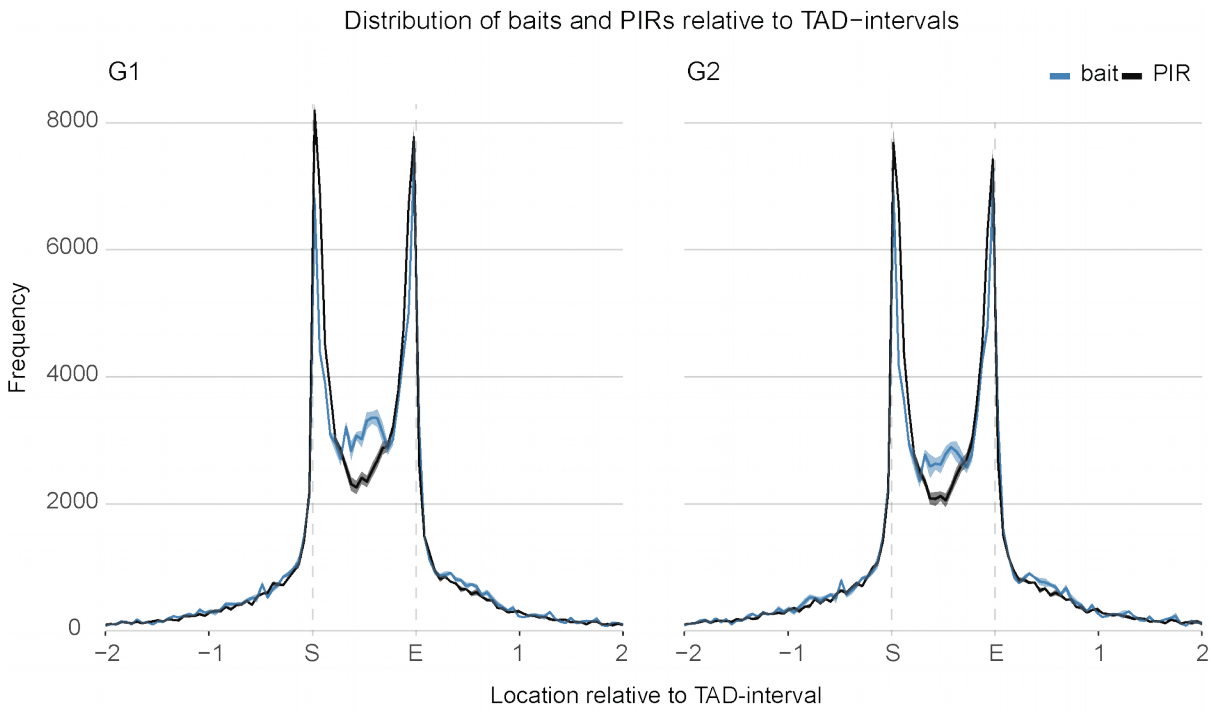

Figure 16. Promoter interactions tend to localise to TAD boundaries. Frequency line plots of RFs involved in promoter interactions, with respect to the most proximal TAD. For each RF, the distance in bp to the most proximal TAD interval is normalised for the length of that TAD interval. The TAD start and end boundaries are shown as ‘S’ and ‘E’ on the x-axis. The blue and grey lines show the mean frequency +/- 1SE over four TAD partitionings (G1R1, G1R2, G2R1, G2R2).

Baits, as well as PIRs, show a clear and strong increase in frequency at TAD boundaries. The frequency of baits and PIRs decreases towards the TAD-centre, although this trend is more pronounced for PIRs than baits. Stringent TADs were used for this analysis. Bait-to-bait interactions were excluded.

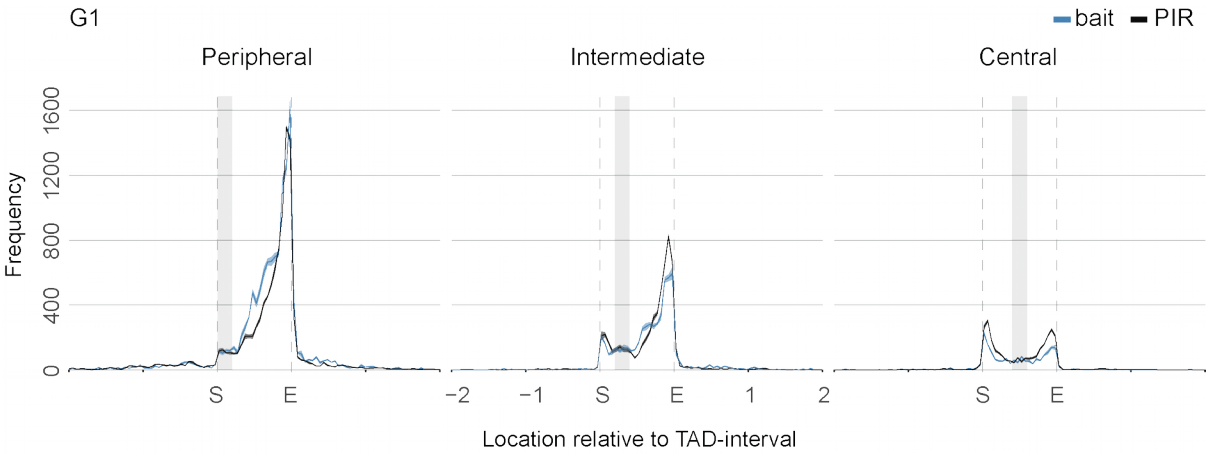

Figure 17. Promoter interactions tend to form between partner RFs at extremes of TADs. Viewpoint window plots of baits and PIRs from G1 PCHiC interactions. The vertical grey bar indicates the position of the viewpoint window: of a pair of interacting RFs, one of the two interacting partners is located within the viewpoint. E.g. the blue line indicates the frequency distribution of baits, the partner PIR of which is located within the window. The line represents the mean of four TAD samples (G1R1, G1R2, G2R1, G2R2), the area around the line indicates +/- 1SE. This figure shows that baits and PIRs have a strong tendency to interact with a partner RF at the opposite TAD-boundary (see peripheral and intermediate viewpoints). Additionally, the frequency of interactions from the central viewpoint is roughly symmetrical, showing that a bait or PIR that is located centrally in a TAD, interacts as readily with either side.

Now that the frequency distributions of baits and PIRs within TADs are known, I move on to evaluating consensus in TAD boundary crossing. To this end I take the union of significant promoter interactions in G1 and G2 and I compare them to the four interphase TAD partitionings individually. I define TAD-boundary crossing as: the interval between two RFs that are involved in an interaction contains at least one TAD boundary (note that this definition does not permit the RFs to intersect with the

TAD boundary and that TAD boundaries are defined as single bp coordinates). I then plot the number of detected TAD boundary crossing events versus the number of TAD partitionings in which crossing is detected (Figure 18). The large proportion of non-crossing interactions (bar at 0) recapitulates that the majority of promoter interactions do not cross TAD boundaries. Additionally, when TAD boundary crossing is detected, it is most commonly detected in all four interphase TAD partitionings (bar at 4). This provides support for the notion that promoter interactions cross TAD boundaries.

To further investigate the TAD boundary crossing consensus between interphase TAD partitionings, I calculated the Jaccard index for boundary crossing events in a pairwise manner. I.e. I calculated the number of times TAD boundary crossing is detected in pairs of TAD partitionings. Then I divided the intersection of these interactions by the union. I performed the same analysis for non-crossing interactions. I visualised the resulting Jaccard indices in two heatmaps in Figure 19.

For inter-TAD interactions, this analysis shows that the Jaccard indices between interphase conditions (G1 versus G2) are lower than within conditions (e.g. G1 versus G1). This reiterates the previous finding (shown in Figure 10) that although similar, the TAD partitionings in G1 and G2 do show some differences. Furthermore, Figure 19 shows that the Jaccard indices for inter-TAD interactions are lower overall than for intra-TAD interactions. This indicates that classification of inter-TAD interactions is more sensitive to differences in detected TAD boundary locations, which in turn suggests that at least one of the RFs involved in inter-TAD interactions is located proximally to a detected TAD boundary.

Using the separate replicate TAD partitionings, I calculate a consensus TAD-crossing Boolean. The Boolean is false if an interphase promoter interaction does not cross any TAD boundary in all replicates (represented by bar 0 in Figure 18). The Boolean is true if an interphase promoter interaction crosses a TAD boundary in all four replicates (represented by bar 4 in Figure 18).

Of the baits with inter-TAD-only promoter interactions, baits with ≥ 2 promoter interactions represent ~59.8%. This shows that a sizeable proportion of baits with inter-TAD-only interactions has multiple interactions of this sort, indicating that these are not spuriously detected promoter interactions. To further analyse the validity of

TAD-boundary crossing interactions, I investigated the consensus between interphase PCHiC interactions (G1 and G2 combined) and the four interphase TAD partitionings. I.e. I raised the following question: How frequently do I detect a TAD boundary that is located in-between a promoter and its PIRs? In other words: if a PCHiC interaction is detected to cross a boundary in G1, how frequently does the same interaction cross a TAD boundary in G2?

This analysis showed that a minority of baits had inter-TAD-only interactions exclusively in G1 or in G2 (287 baits in either, as opposed to 816 “consistent inter-TAD-only” baits, see Figure 20). This indicates that inter-TAD-only promoter interactions are generally stable throughout interphase, though not exclusively. This result is further examined in the Discussion (section 5.1.4).

I show an example of an “inter-TAD-only” bait (*TLR10* promoter region) in Figure 21. For visualisation purposes, I selected a very high-resolution Hi-C contact map comprised of 10 combined Hi-C libraries from PGP1f cells (Nir *et al.*, 2018). The lower-resolution Hi-C data in HeLa cells, are comparable in the region presented in this figure (see inset in Figure 21). TAD boundary detection resulted in three different start boundary positions and one end position for the TAD immediately upstream of *TLR10*, with the baited *TLR10* promoter itself mapping outside of the TAD on the right. In contrast, all *TLR10* PIRs are located inside of the TAD, and proximally to the three detected TAD start boundaries. This figure illustrates that a bait that is not located within a TAD can still interact with regions within that TAD. Moreover, even though the bait fragment is located outside of the TAD, the fact that all PIRs map to the distal TAD boundaries suggests that the TAD structure is nonetheless relevant for these interactions. Finally, the positioning of the bait beyond the TAD, and the lack of clear Hi-C interactions from the bait region to its PIRs, highlights that PCHiC is capable of detecting interactions with higher sensitivity than Hi-C alone.

I next sought to identify where all RFs with inter-TAD-only interactions are located relative to TAD boundaries. However, the symmetry in the RF frequency within TADs (Figure 16) cannot unambiguously inform on the location of an interacting RF relative to its interacting partner. Furthermore, the viewpoint window approach (Figure 17) excludes RFs that are located beyond TAD boundaries. To address these limitations, I adapted the viewpoint window analysis to “half TAD+ windows” that cover 50% of

the TAD length plus the region beyond the boundary. This analysis reveals the position of inter-TAD-only baits relative to first half of the most proximal TAD, which shows that many inter-TAD-only baits are located adjacently to a detected TAD boundary but just outside the TAD (Figure 22). Furthermore, the PIRs of these baits are strongly overrepresented at the opposite TAD boundary. Conversely, intra-TAD-only baits approximate a uniform distribution within TADs and their PIRs are most frequently located at the TAD boundaries. This analysis demonstrated, similarly to the above example of *TLR10*, that baits with inter-TAD-only interactions were most commonly located just outside the detected TAD intervals. This indicates that in many of these cases, inaccuracies in the detection of TAD boundary locations result in apparent TAD boundary crossing. Therefore it is likely that many of these detected TAD boundary crossing events are misclassified. Regardless of the relation to TAD boundary crossing, the detected baits with inter-TAD-only interactions exhibit a strong tendency to interact at the opposite TAD boundary. In contrast to baits with intra-TAD-only interactions, which interact at either side of the TAD. Taken together, this suggests that a subset of the inter-TAD-only baits are involved in boundary to boundary looping interactions and that yet another subset comprise of true TAD boundary crossing interactions.

To de-convolve true inter-TAD-only baits from baits at TAD boundaries which are potentially misclassified as inter-TAD-only, I exclude baits that are located within 20% of the TADs length at either side of the boundary. This results in 229 baits with high-confidence inter-TAD-only interactions. Figure 23 shows the corresponding frequency distribution (as a ‘half TAD+’ graph). Since these promoter interactions appear to defy insulation by TAD boundaries (at least to an extent), I next ask the question if this relates to TF binding, histone modifications or transcriptional activity.

To this end, in the next section, I describe functional differences between inter-TAD-only and intra-TAD-only baits and PIRs, making use of the stringent set of baits with inter-TAD-only interactions.

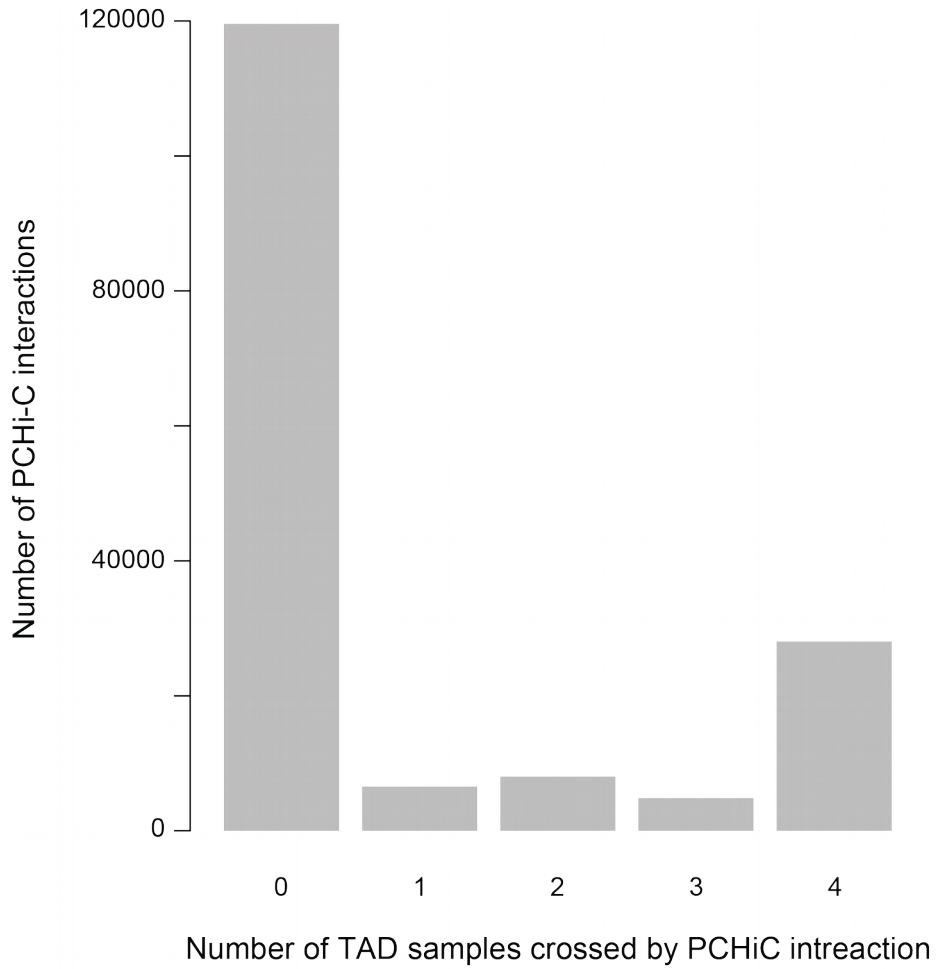

Figure 18. The frequency of interphase promoter interactions crossing a TAD boundary in 0 to 4 TAD partitionings. This figure shows that there is general agreement between TAD replicates in terms of promoter interactions crossing TAD boundaries: bars at 0 and 4 are the largest. I.e. when a TAD boundary is located in-between the bait and PIR of an interphase promoter interaction, this boundary tends to be detected in all four TAD partitionings (G1R1, G1R2, G2R1 and G2R2). Note that the union of significant G1 and G2 promoter interactions are used. Bar heights: 119576, 6555, 8033, 4848, 28037

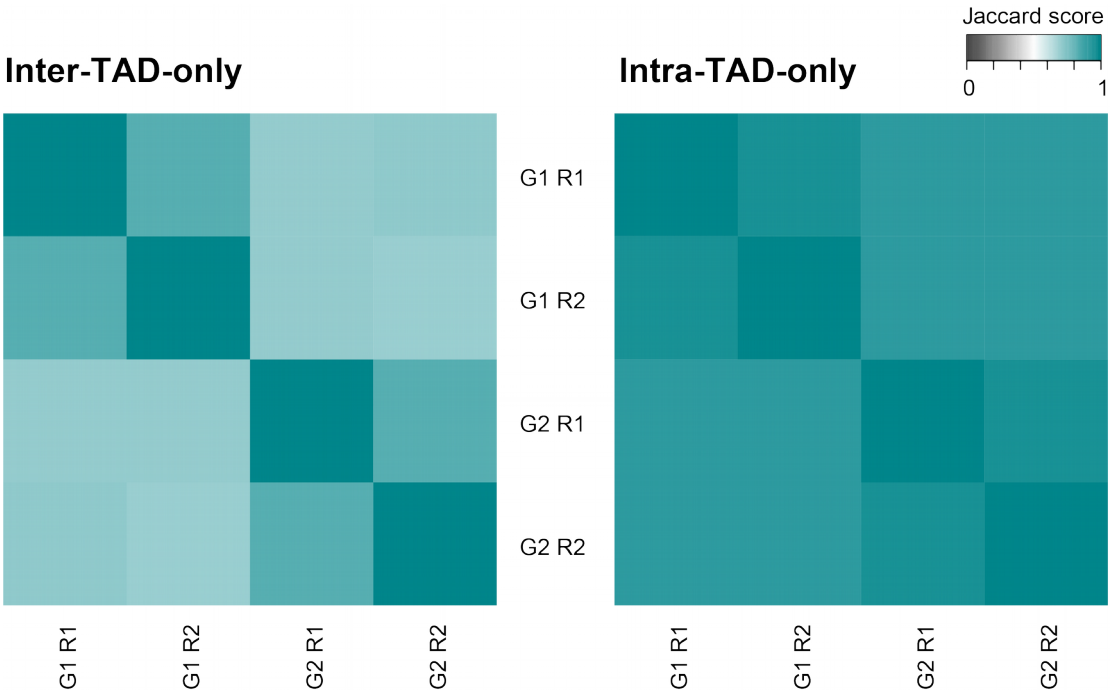

Figure 19. Jaccard indices showing pairwise sample concurrence in TAD-boundary crossing of interphase promoter interactions. The jaccard index is calculated by dividing the number of shared TAD-boundary crossing events for a pair of TAD samples by the total number of TAD-boundary crossing events for the same pair of TAD samples. Inter-TAD-only: G1 replicates show a jaccard score of ~0.822. G2 replicates show a jaccard score of ~0.820. G1 and G2 samples show jaccard scores within the range [0.698, 0.710]. Intra-TAD-only: G1 replicates show a jaccard score of roughly 0.945. G2 replicates show a jaccard score of roughly 0.945. G1 and G2 samples show jaccard scores within the range [0.904, 0.907].

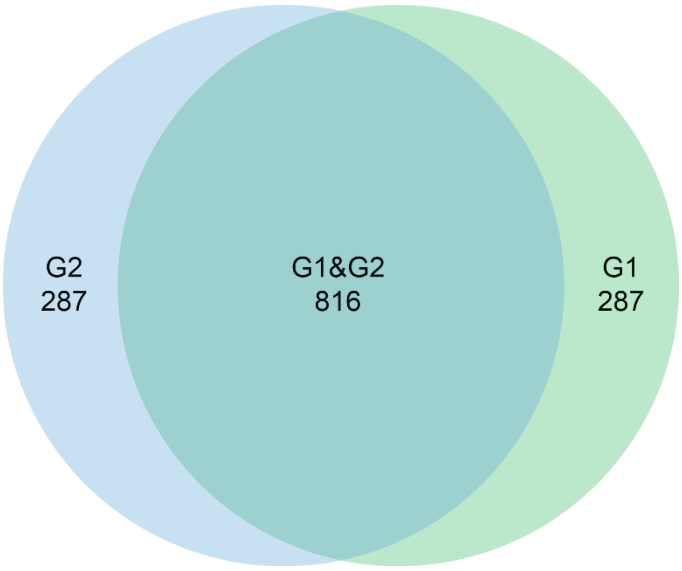

Figure 20. The majority of promoters with inter-TAD-only interactions are stable between G1 and G2. Venn diagram of baits with inter-TAD-only interactions in G1 and G2. Baits with inter-TAD-only interactions show a Jaccard index of ~0.587 between G1 and G2. Note that TAD-boundary crossing requires consensus between replicates (e.g. a TAD boundary crossing interaction in G1, crosses a TAD boundary in G1R1 and G1R2).

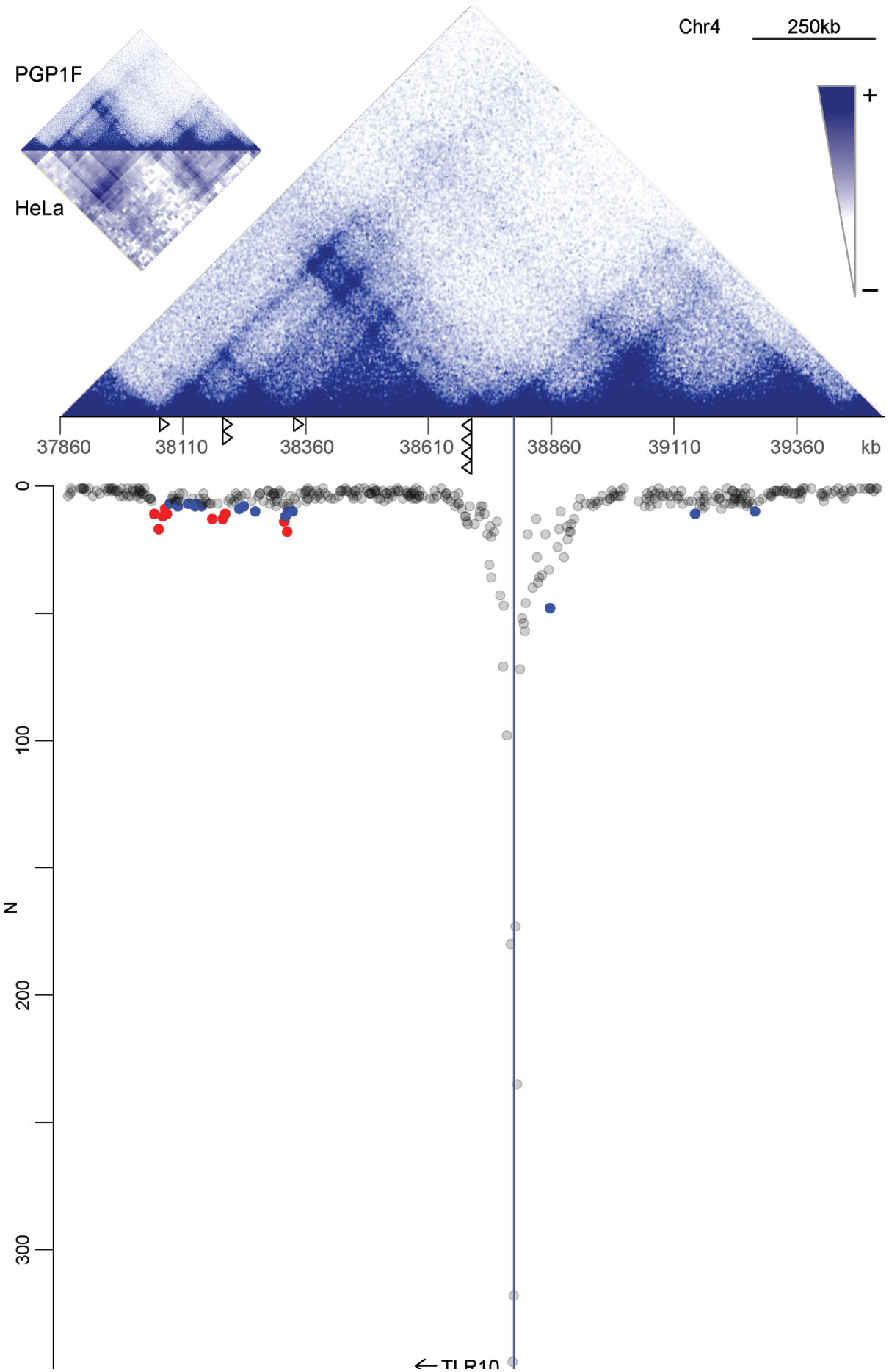
PGP1F
HeLa
Chr4
250kb
+
−
37860
38110
38360
38610
38860
39110
39360
kb
0
100
200
300
N
←TLR10

Figure 21. Example of a bait with inter-TAD-only interactions. A human fibroblast PGP1f cell line derived high-resolution Hi-C heat map is shown at the top (Nir *et al.*, 2018) (chr4: [37860000, 39540000]). This Hi-C contact map is comparable to the HeLa contact maps of the same region (see top-left inset). However, the resolution is much higher (5kb vs 40kb). Below the Hi-C contact map, black triangles show the coordinates of detected TAD boundaries in HeLa interphase samples G1 and G2 (2 replicates each). The black triangles point inwards to differentiate TAD-start and TAD-end boundaries. The blue line shows the location of the baited RF, in this instance the bait contains the promoter for the gene *TLR10*. Red circles on the PCHiC plot (bottom) correspond to PIRs with a significant CHiCAGO score (≥ 5). The baited RF is located ~80kb beyond the TAD-end boundary but its significantly interacting RFs are located proximally to the detected TAD-start boundaries. The heat map was visualised using Juicebox (Durand *et al.*, 2016).

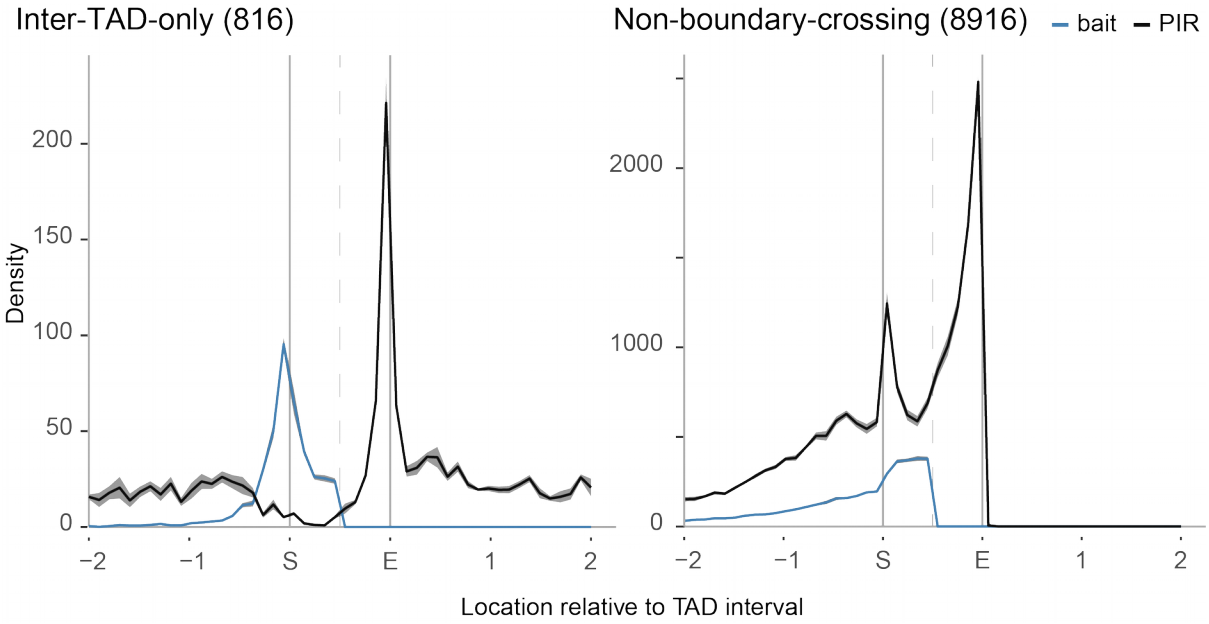

Figure 22. Frequency distribution of baits (blue) and PIRs (grey) from inter-TAD-only and non-boundary-crossing-only promoter interactions. The diagrams show the frequency normalised to the most proximal TAD of the bait. The shaded area represents +/- 1SE over 4 TAD partitionings. The interactions are mirrored along the vertical axis at the TAD-centre (dashed grey line). Baits of inter-TAD-only promoter

interactions are most commonly located proximally to a TAD boundary, but outside of the TAD interval. The PIRs of these baits are most frequently located at the opposite TAD boundary and within the TAD interval. Baits of non-boundary-crossing-only promoter interactions appear to be approximately uniformly distributed within TADs. The PIRs of these baits are most frequently located at the opposite TAD boundary of the bait, however, an increase in frequency also appears at the TAD boundary on the same side as the bait. Furthermore, the PIRs of non-boundary-crossing-only baits that are not located within any TAD, show a slight increase in frequency proximally to the nearest TAD-boundary of the bait.

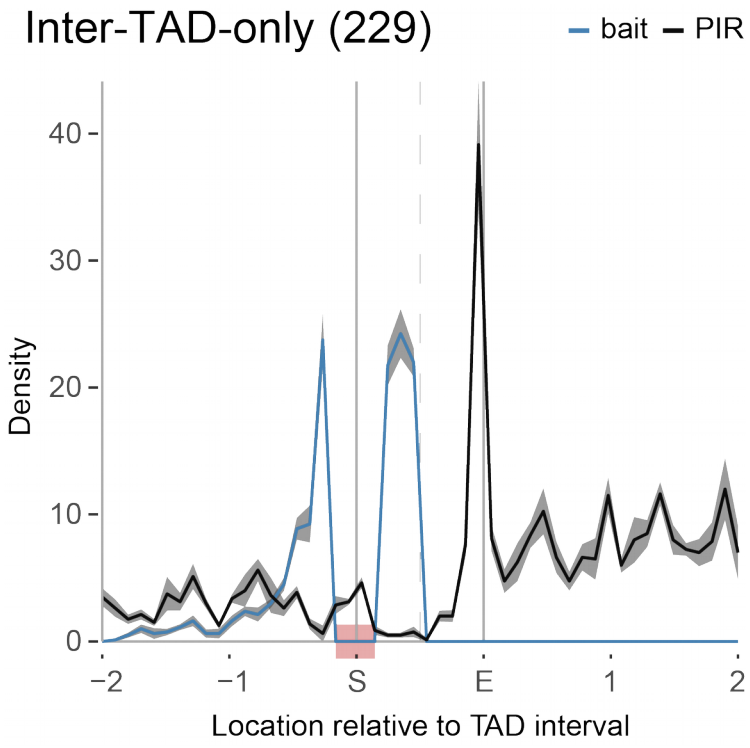

Figure 23. Frequency distribution of baits (blue) and PIRs (grey) from inter-TAD-only promoter interactions while excluding boundary-proximal baits (red area). TAD boundary proximal baits have been excluded (red shaded area; +/- 20% TAD length). The diagram shows the frequency normalised to the most proximal TAD of the bait. The shaded area represents +/- 1SE over 4 TAD partitionings. The interactions are mirrored along the vertical axis at the TAD-centre. The PIRs of these baits are frequently located at the opposite TAD boundary. However, PIRs

may also be located beyond the boundaries of the bait-proximal TAD (black line outside the S-E interval).

3.5 Functional properties of inter-TAD-only and intra-TAD-only interactions

To investigate the functional differences between RFs involved in inter-TAD-only and intra-TAD-only interactions, I performed three analyses: Gene Ontology enrichment analysis, RNA-Seq data analysis, and ChIP-Seq data analysis. The GO and RNA-Seq analyses were performed on the genes with promoters within the baits involved in inter-TAD-only or intra-TAD-only interactions.

The GO enrichment analysis yielded no significant results (using the two online tools GOrilla and LAGO) on genes from the non-stringent set of baits as well as the stringent set of baits. This indicates that neither sets of genes (with inter-TAD-only or with intra-TAD-only promoter interactions) show clear similarities in terms of GO annotations. Note that this analysis includes all three groups of the gene ontology: molecular function, biological process and cellular location.

I next contrasted transcriptional activity of genes with promoter interactions in the inter-TAD-only group to the intra-TAD-only group, using RNA-Seq data (produced by Gordana Wutz). Figure 24 shows the distributions of gene-length-normalised, log-transformed read counts per gene. This shows that baits with inter-TAD-only interactions had lower transcript abundance than baits with intra-TAD-only interactions. A Mann-Whitney U-test reports a significant difference with a p-value of ~0.00008238. However, when I repeated this analysis on genes that correspond to the stringent set of baits (described in section 3.4). This shows no discernible difference in transcript levels (Figure 25), and the Mann-Whitney U-test p-value is insignificant (0.8564). Since the stringent set of interactions are less likely to contain false positives, these results suggest that genes with inter-TAD-only promoter interactions and those with intra-TAD-only promoter interactions do not show clear differences in transcriptional activity. Furthermore, the differences that are found using the non-stringent definition of inter/intra-TAD-only interactions suggest that the TAD boundary related properties of these interactions may be involved in the observed difference in transcript levels.

To investigate whether the difference in transcript levels (of genes with non-stringent inter/intra-TAD-only interactions) was underpinned by the number of promoter interactions per bait, I stratified baits by their number of interactions and then

visualised the transcript levels as violin plots. Figure 26 shows no clear relation between the number of interactions per bait and transcriptional activity. Table 6 shows the number of baits stratified by their numbers of interactions. Even though the transcript levels vary between strata, the median value is lower in the inter-TAD-only set compared to the intra-TAD-only set, regardless of the number of promoter interactions. Therefore it is unlikely that the observed transcriptional suppression is imparted by the PIRs. This suggests that the lower transcript level of genes with inter-TAD-only promoter interactions is driven by a different mechanism, perhaps directly at the promoter regions. Given that the non-stringent baits with inter-TAD-only promoter interactions are frequently located proximally to TAD boundaries, this suggests that TAD boundary regions themselves may suppress transcription. Collectively, the observed difference in transcript levels between inter-TAD-only and intra-TAD-only promoter interactions is likely not driven by promoter interactions. It remains to be seen which mechanism is responsible.

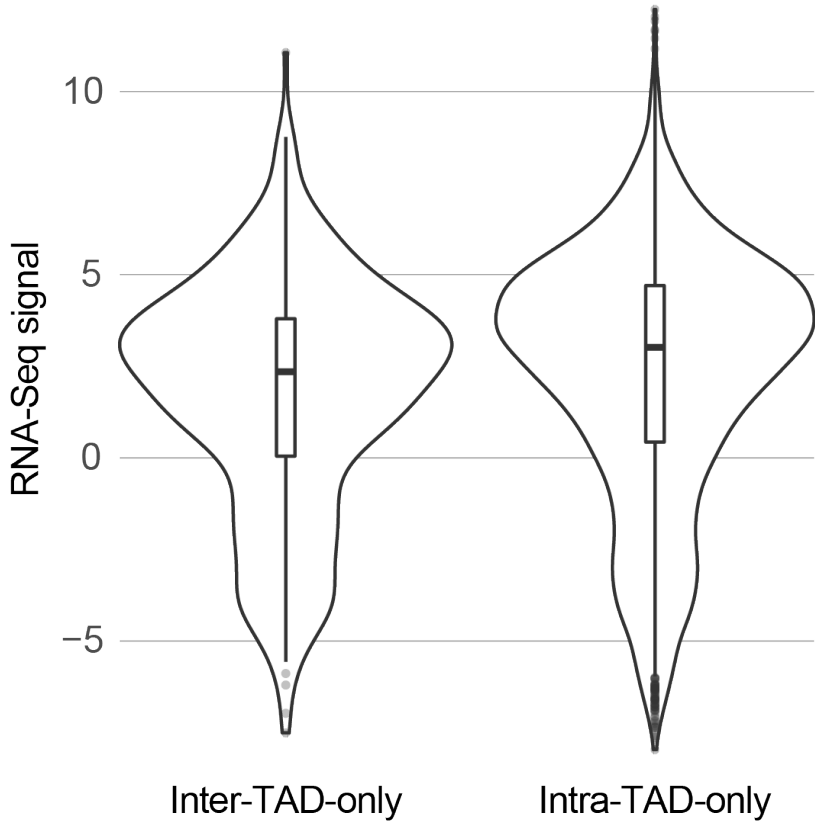

Figure 24. Promoters with inter-TAD-only promoter interactions show lower transcript levels than promoters with intra-TAD-only interactions. Gene-length corrected, log-transformed RNA-Seq read

counts (y-axis: expr) mapped to genes with promoters with inter-TAD-only and intra-TAD-only interactions. I excluded genes with no RNA-Seq reads, and I took the mean over two replicates. N inter-TAD-only promoters: 354. N intra-TAD-only promoters: 4764. Genes in the “all-crossing” category show lower RNA-Seq signal than genes in the “none-crossing” category. Mann-Whitney p-value ~8.2*10-5.

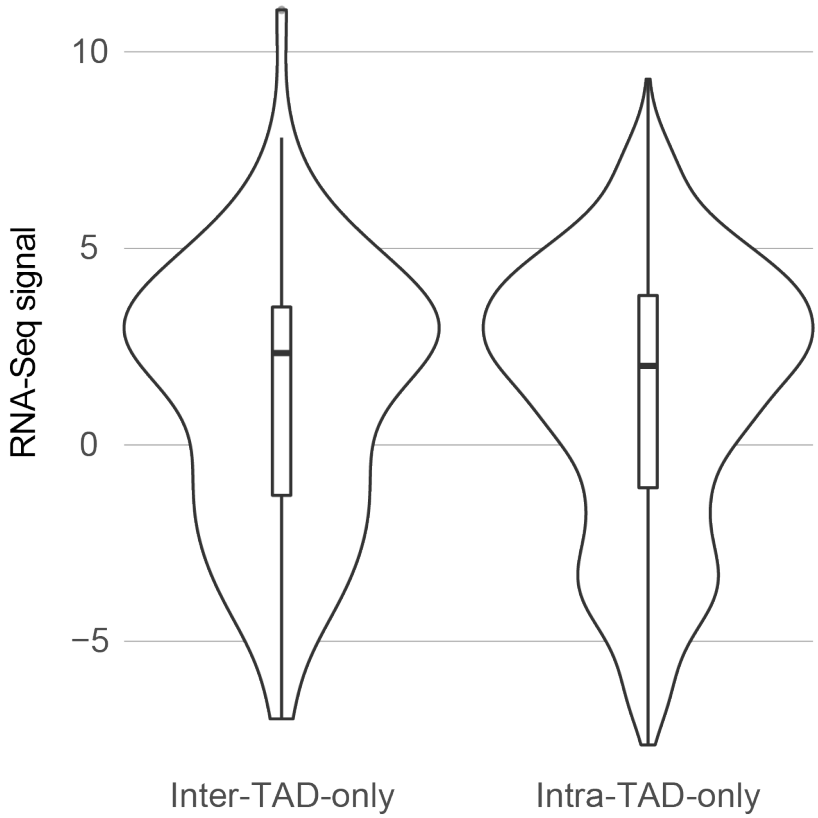

Figure 25. Non-boundary-proximal promoters with inter-TAD-only promoter interactions show similar transcript levels to promoters with intra-TAD-only interactions. Gene-length corrected, log-transformed RNA-Seq read counts (y-axis: expr) mapped to genes with promoters with inter-TAD-only and intra-TAD-only interactions. I excluded genes with no RNA-Seq reads, and I took the mean over two replicates. Genes in the “all-crossing” category show similar RNA-Seq signal to genes in the “none-crossing” category. Mann-Whitney p-value: 0.8564

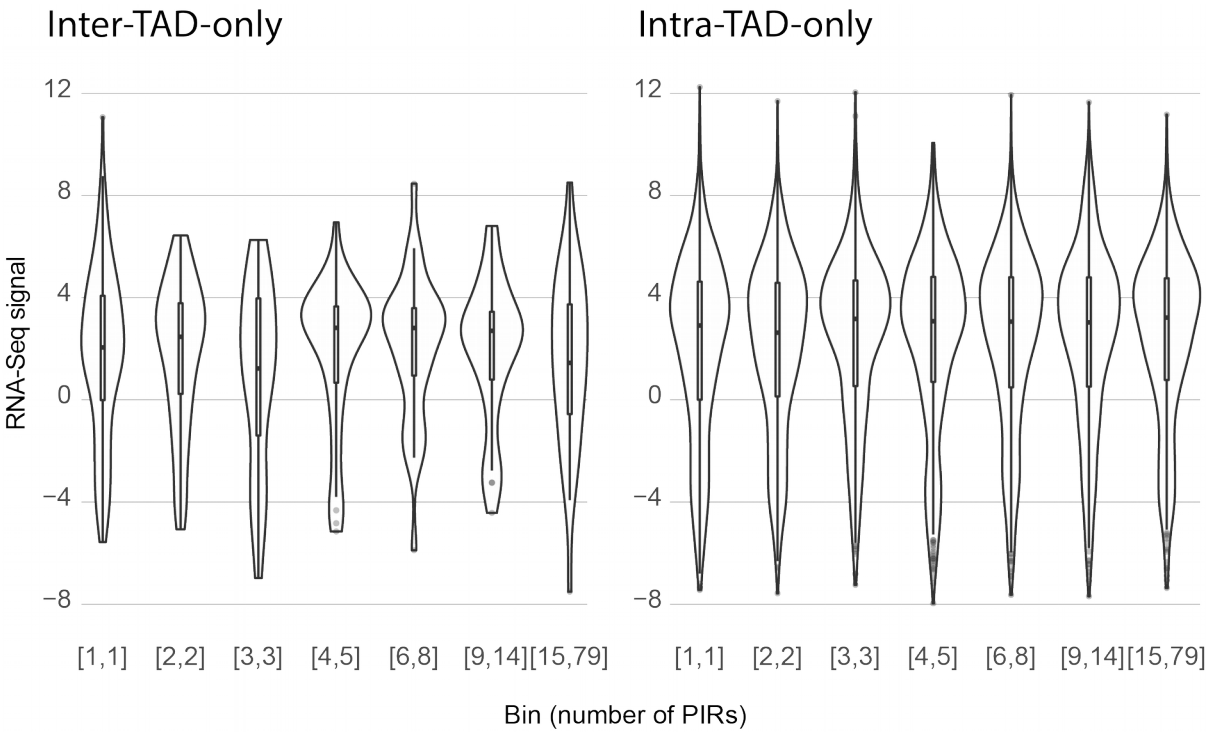

Figure 26. Lower transcript levels for genes with inter-TAD-only promoter interactions appears unrelated to the number of interactions. Gene-length corrected, log-transformed RNA-Seq reads mapped to genes with promoters with inter-TAD-only (left) and intra-TAD-only (right) interactions, stratified by number of promoter interactions. No clear relation can be seen between the number of promoter interactions and transcript levels. Note that non-stringent inter-TAD-only promoter interactions were used in this analysis.

Table 6. Number of interactions per bin. Intra-TAD-only interactions are more common than inter-TAD-only interactions. However, baits with inter-TAD-only interactions do frequently interact with multiple PIRs across a TAD boundary.

	Bin (number of interactions)						
	[1,1]	[2,2]	[3,3]	[4,5]	[6,8]	[9,14]	[15,78]
Intra-TAD-only	976	652	549	671	712	709	656
Inter-TAD-only	134	62	35	42	33	31	31

Next, I sought to investigate which DNA-binding proteins and histone modifications associated with RFs involved in inter-TAD-only versus intra-TAD-only promoter interactions. To this end I devised a strategy using publicly available ChIP-Seq data as well as those from my collaborators (totalling data on 51 DNA binding proteins and 11 histone targets, after quality control), which I also employed to address ChIP-Seq signal enrichment at RFs in further analyses presented later in the text. This approach is not based on discrete binding peaks detected at RFs with conventional ChIP-Seq peak-calling software, but rather, integrates the ChIP-Seq signal over entire RFs and expresses this enrichment in the form of statistically meaningful scores (see Methods 2.11). To detect significant associations of individual factors and histone modifications with RFs, I performed LASSO logistic regression using the per-RF ChIP-Seq scores as independent variables. The binary dependent variable categorizes an RF as being involved in inter-TAD-only or intra-TAD-only interactions. To account for the effects of the CHiCAGO score and interaction distance, I also included these values in the regressions. I performed these analyses for inter-TAD-only baits in G1 and G2 separately. Note also that I performed separate analyses on the non-stringent set of inter/intra-TAD-only RFs (including boundary proximal RFs) and on the stringent set of inter/intra-TAD-only RFs (Figure 27 and Figure 28 respectively). Figure 27 shows the LASSO logistic regression coefficients per penalty factor lambda for baits and PIRs from the full set of inter/intra-TAD-only interactions. Figure 28 the lasso logistic regression coefficients for the stringent (boundary proximity excluded) set. A segmented vertical line shows the minimal optimal

lambda, corresponding to a level of shrinkage that favours fewer independent variables while remaining within 1SE of the optimal lambda. The significance of the regression coefficients was calculated for this “minimal-optimal” level, with the significant predictors shown in colour. The sign of the regression coefficients corresponds to their association with RFs involved in inter-TAD-only (positive) or intra-TAD-only (negative) promoter interactions, respectively. Interaction distance was detected as a significant predictor and had a positive sign in both G1 and G2, indicating its association with inter-TAD-only RFs. This is expected, given that longer-range interactions are more likely to cross TAD boundaries. In contrast, the CHiCAGO score significantly associated with intra-TAD-only interactions, showing that such interactions tend to show stronger signal overall. Notably, the analyses also revealed the preferential association of a number of ChIP-Seq targets (summarised in Table 7 and Table 8).

This analysis has revealed that MYBL2, PWWP2A, and GATAD1 are associated with inter-TAD-only promoter interactions in the non-stringent set. The stringent set of inter-TAD-only promoter interactions associated with SGF29 and ZZZ3.

To investigate common properties of the detected factors, I performed gene ontology enrichment analysis using LAGO (Boyle *et al.*, 2004) (https://go.princeton.edu/cgi-bin/LAGO). This revealed that the detected factors at RFs with inter-TAD-only promoter interactions, with the exception of PWWP2A, share the high-level GO term “transcriptional regulator activity” (see Table 9). In addition, SGF29 and ZZZ3 shared the terms “histone acetyltransferase process” and “transcriptional activator complex”. Taken together, these findings suggest that inter-TAD-only promoter interactions may include promoter-enhancer interactions. This is surprising, since the RNA-Seq analysis presented above suggests that transcription of genes with promoters that interact inter-TAD-only is potentially lower than intra-TAD-only. This is further examines in the Discussion section 5.1.3.

Of the ChIP-Seq factors that were detected at RFs involved in intra-TAD-only promoter interactions, BRD4, HJURP and YY1 shared the high-level GO term “chromatin organisation”. This suggests that these factors may be involved in shaping the chromatin landscape, perhaps in such a way that enables inter-TAD promoter interactions. This is supported by the recent discovery of YY1 as a

structural regulator of promoter-enhancer loops (Beagan *et al.*, 2017; Weintraub *et al.*, 2017)

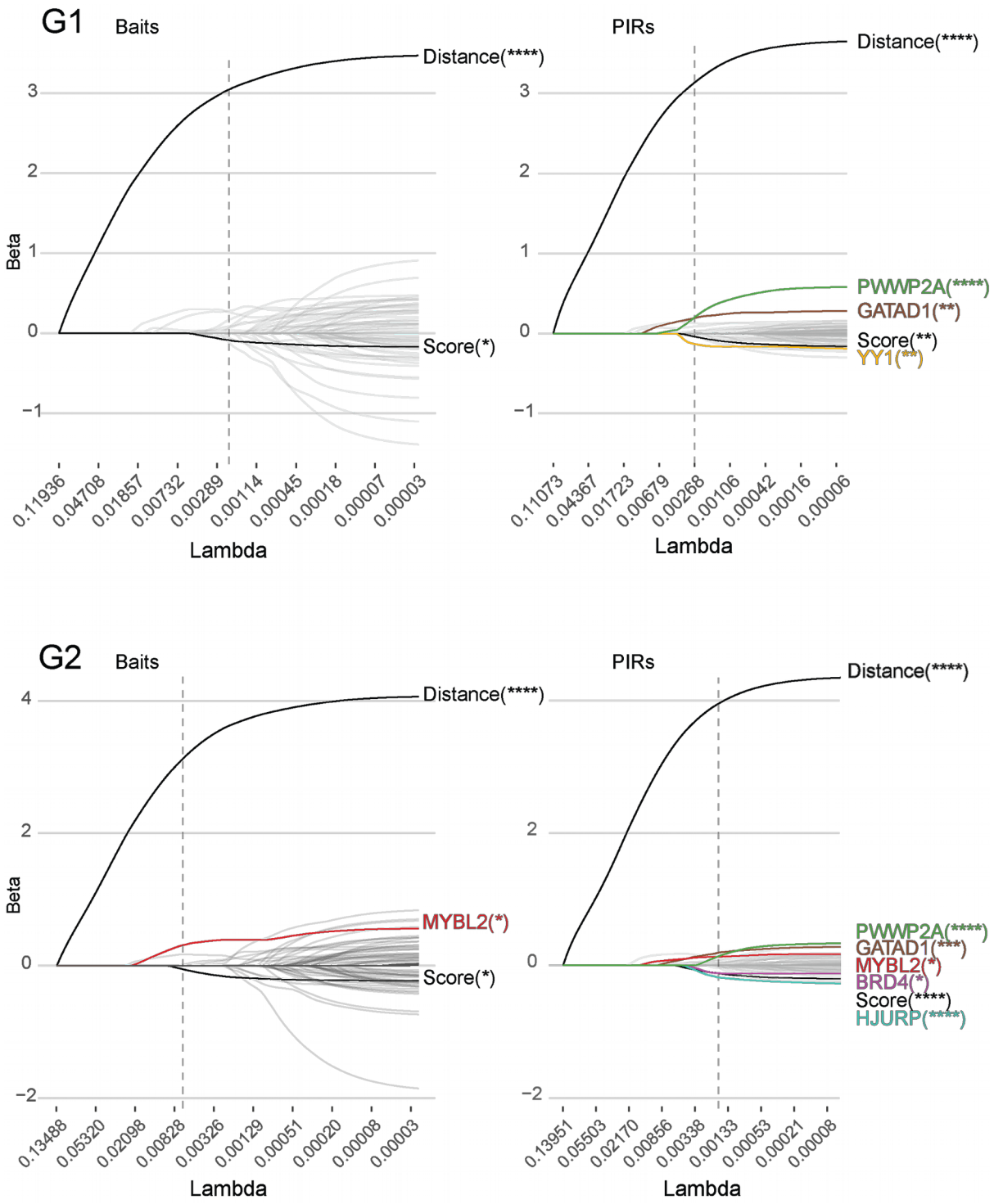

Figure 27. LASSO logistic regression reveals TAD-boundary-crossing-specific PIR properties. LASSO regularisation paths of ChIP-Seq signal at baits (left) and PIRs (right) of promoter interactions in G1 (top) and G2 (bottom). The dependent variable in both cases is “inter-

TAD-only" vs "intra-TAD-only". The independent variables consist of pre-processed ChIP-Seq scores per RF. Additionally, the independent variables include the mean log10 absolute genomic distance between the bait and the PIR and the mean CHiCAGO score of the interactions (analyses were also performed using the minimum and maximum distance and score, yielding comparable results). The vertical dotted lines represents the lasso regression lambda value that is within 1SE from the optimum, biased towards a minimal number of non-zero beta's. Independent variables with a significant beta at the lambda 1SE mark are given a colour. The promoter interaction distance and CHiCAGO score are significant predictors of promoters with inter-TAD-only and intra-TAD-only interactions respectively. Lastly, in G2 only, MYBL2 is a significant predictor of inter-TAD-only promoter interactions.

Table 7. Significant ChIP-Seq signals at baits and PIRs of inter-TAD-only versus intra-TAD only interactions in G1 and G2. MYBL2, PWWP2A and GATAD1 correlate with inter-TAD-only interactions. YY1, BRD4 and HJURP correlate with intra-TAD-only interactions.

	ChIP-Seq target	G1	G2
Inter-TAD-only	MYBL2	-	Bait + PIR
	PWWP2A	PIR	PIR
	GATAD1	PIR	PIR
Intra-TAD-only	YY1	PIR	-
	BRD4	-	PIR
	HJURP	-	PIR

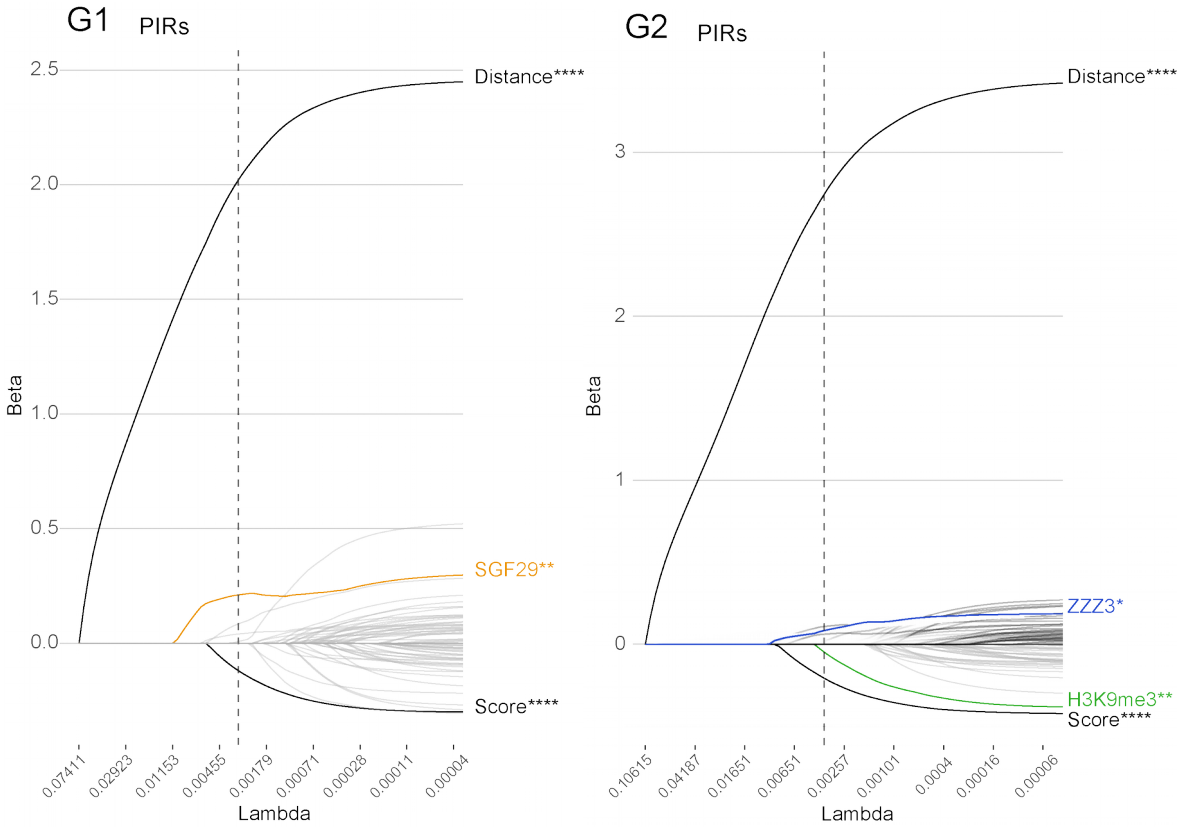

Figure 28. LASSO logistic regression reveals stringent TAD-boundary-crossing-specific PIR properties. LASSO regularisation paths of ChIP-Seq signal at PIRs of promoter interactions in G1 (left) and G2 (right). The dependent variable in both cases is “inter-TAD-only” vs “intra-TAD-only”, however in this analysis baits are excluded when they are located proximally to a TAD boundary. The independent variables were the same as the previous lasso logistic regression analysis. Chromatin organisation related factors SGF29 and ZZZ3 are identified at PIRs of stringent inter-TAD-only interactions. The repressive mark H3K9me3 is identified at PIRs of stringent intra-TAD-only interactions. The results on baits yielded no significant results (other than distance and score) and were therefore excluded from this figure.

Table 8. Significant ChIP-Seq signals at PIRs of inter-TAD-only versus intra-TAD-only interactions in G1 and G2. SGF29 and ZZZ3 correlate with inter-TAD-only interactions. H3K9me3 correlates with intra-TAD-only interactions.

	ChIP-Seq target	G1	G2
Inter-TAD-only	SGF29	PIR	-
	ZZZ3	-	PIR
Intra-TAD-only	H3K9me3	-	PIR

Table 9. Gene ontology enrichment analysis results for inter-TAD-only ChIP-Seq targets. SGF29 and ZZZ3 share histone 4 acetyltransferase activity, which returns 5 related GO-terms. GATAD1, MYBL2, SGF29 and ZZZ3 share the high-level GO term: transcription regulator activity.

GO ID	Description	FDR	ChIP-Seq IDs
0005671	Ada2/Gcn5/Ada3 transcription activator complex	0.000138385	SGF29, ZZZ3
1902562	H4 histone acetyltransferase complex	0.00260283	SGF29, ZZZ3
0000123	histone acetyltransferase complex	0.00841193	SGF29, ZZZ3
0031248	protein acetyltransferase complex	0.0111891	SGF29, ZZZ3
1902493	acetyltransferase complex	0.0111891	SGF29, ZZZ3
0140110	transcription regulator activity	0.0428923	GATAD1, MYBL2, SGF29, ZZZ3

Table 10. Gene ontology enrichment analysis results for intra-TAD-only ChIP-Seq targets (non-histone marks). BRD4, HJURP and YY1 share the high-level GO term: chromosome organization.

GO ID	Description	FDR	ChIP-Seq IDs
0051276	chromosome organization	0.0049948	BRD4, HJURP, YY1

4 Investigating the effects of architectural protein perturbations on promoter interactions and transcription

In the previous chapter I showed that promoter interactions are constrained by TAD boundaries, though but not fully contained within them. I show evidence that suggests that TAD boundary crossing may occur between promoters and enhancers.

Architectural proteins such as CTCF and cohesin complex, play key roles in TAD definition and maintenance, and may also be involved in facilitating promoter-enhancer interactions (Schoenfelder *et al.*, 2019). Therefore, I investigated the effects of rapid architectural protein depletion on promoter interactions and gene expression. To this end, I analysed PCHiC data from HeLa cells where CTCF and the cohesin subunit SCC1 were independently rapidly depleted, in addition to cells where the cohesin unloading factor Wapl (and its cofactors PDS5A and PDS5B) are knocked down. I detect promoter interaction loss, maintenance, and gain, and I use these rewiring properties to categorise promoter interactions. I investigate these categories in terms of interaction distance, protein binding, histone modifications, nascent transcriptional response, and DNA accessibility. I found that down-regulated genes tend to lose promoter interactions, transcriptionally stable genes tend to maintain promoter interactions, and up-regulated genes tend to maintain and gain promoter interactions. Additionally, I found that DNA accessibility at PIRs correlates with transcriptional change upon SCC1 depletion, in the context of promoter interaction rewiring. Lastly, I describe validation experiments to test our findings, in addition to the preliminary results that these experiments have yielded.

4.1 Analysis of promoter interaction rewiring upon perturbation of architectural proteins

To investigate the genome-wide response of promoter interactions upon perturbation of cohesin, CTCF and WAPL, I first sought to relate the corresponding promoter interaction landscapes to those of cells in cell cycle stages G1, G2 and mitosis. The reasoning being that there are indications that TADs are dynamic across the cell cycle (Nagano *et al.*, 2017), in addition to the observation that TAD organisation is naturally absent in Mitosis (Naumova *et al.*, 2013). I analysed PCHiC data from HeLa cells where CTCF and the cohesin subunit SCC1 were independently rapidly depleted, in addition to cells where the cohesin unloading factor Wapl (and its cofactors PDS5A and PDS5B) are knocked down. I partitioned PCHiC data based on the interaction score using k-means clustering (see Methods 2.5). Briefly, I arrived at a partitioning with 13 clusters by minimising the variance within clusters while simultaneously minimising the number of clusters with similar centroid values. I observed that cohesin depletion shows the most dramatic rewiring response. The clustered CHiCAGO scores (igure 29) provide evidence for three distinct effects: promoter interaction loss, maintenance and gain upon SCC1 and CTCF depletion, WAPL perturbation and in mitosis (see Table 11). Although SCC1 and CTCF depletion results in many promoter interactions being lost, it is also clear that a large proportion of promoter interactions appear unaffected. Furthermore, upon SCC1 and CTCF depletion novel promoter interactions are detected.

Contrasting the SCC1 and CTCF depleted samples to promoter interactions from the more condensed chromatin landscapes of mitotic and WAPL/PDS5A/PDS5B inhibited cells reveals distinct features of promoter interactions in the context of architectural protein depletion: a large number of promoter interactions appear to be cohesin and CTCF independent, even though these interactions are absent in the more condensed landscape (cluster A; 48,572 interactions).

Conversely, many promoter interactions that are lost in the condensed chromatin conditions (WAPL knockdown and Mitosis) are retained upon depletion of cohesin/CTCF (cluster B for WAPL KD and mitosis, clusters C, D and F for mitosis only). It should be noted that the treatment time for knockdown is longer than the rapid depletion conditions (see Methods 2.18), so I cannot exclude the possibility

that the WAPL knockdown data reflect secondary effects rather than the direct effects of perturbation.

Clusters I an M show WAPL knockdown and mitosis specific promoter interactions, respectively, suggesting that promoter interactions in condensed chromatin compose a distinct and separate landscape from cohesin and CTCF depleted conditions. Taken together, these findings show that cohesin and CTCF depletion results in promoter interaction rewiring that is distinct from unperturbed interphase organisation as well as the more condensed chromatin landscapes of Mitosis and the vermicelli phenotype that is the result of triple knockdown of the cohesin release factors WAPL/PDS5A/PDS5B.

Clusters H and L are puzzling, since the interphase promoter interactions from cluster H are not present in the control samples SCC1 -Auxin and CTCF -Auxin, and because the promoter interactions in cluster L occur only in the control condition SCC1 -Auxin. One possible explanation for the interactions in cluster H is that the fusion of target proteins SCC1 and CTCF to the IAA1771 degron partially impairs protein function. Alternatively, the degron system may show low levels of activity in the absence of Auxin. A possible explanation for the interactions in cluster L is that the interactions may have arisen clonally, reflecting distinct DNA organisation of the single common ancestor cell. Altogether, the promoter interactions in clusters H and L represent form small proportion of all detected interactions. This suggests that the extent of the effects that underlie these unexpected promoter interactions is minimal. Promoter interactions in clusters H and L are discarded from further analysis.

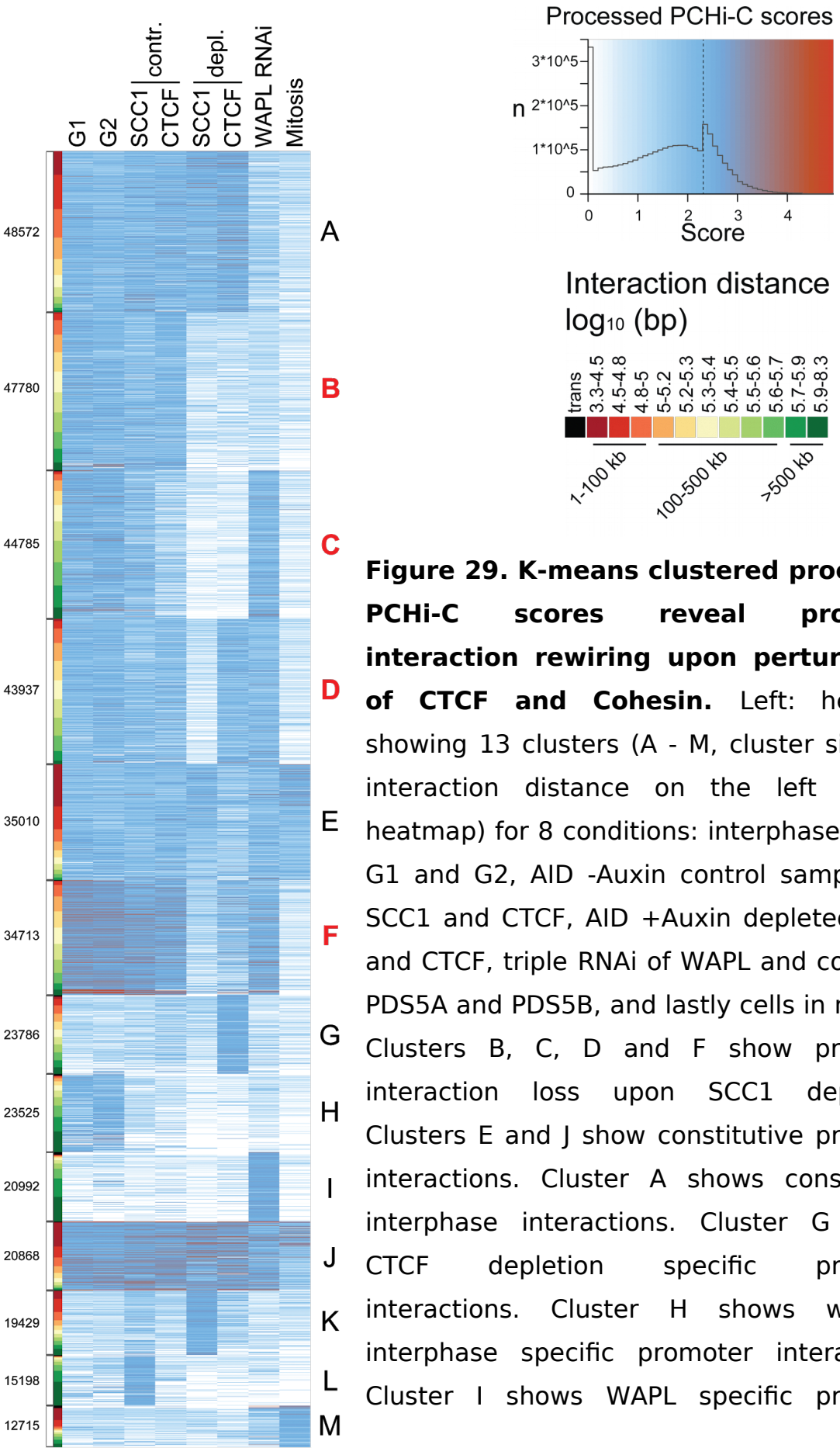

Figure 29. K-means clustered processed PCHi-C scores reveal promoter interaction rewiring upon perturbation of CTCF and Cohesin. Left: heatmap showing 13 clusters (A - M, cluster size and interaction distance on the left of the heatmap) for 8 conditions: interphase stages G1 and G2, AID -Auxin control samples for SCC1 and CTCF, AID +Auxin depleted SCC1 and CTCF, triple RNAi of WAPL and cofactors PDS5A and PDS5B, and lastly cells in mitosis. Clusters B, C, D and F show promoter interaction loss upon SCC1 depletion. Clusters E and J show constitutive promoter interactions. Cluster A shows constitutive interphase interactions. Cluster G shows CTCF depletion specific promoter interactions. Cluster H shows wildtype interphase specific promoter interactions. Cluster I shows WAPL specific promoter

interactions. Cluster K shows SCC1 depletion specific promoter interactions. Cluster L shows AID -Auxin SCC1 control specific promoter interactions. Cluster M shows mitosis specific promoter interactions.

Table 11. Promoter interaction rewiring classification according to the clustered CHiCAGO score response upon various perturbations. Classification was performed relative to promoter interaction organization in the control stages. *Cluster H shows decreases CHiCAGO scores between interphase and control conditions, therefore I excluded these promoter interactions in this table. Cluster L shows increases CHiCAGO scores between G1, G2 and CTCF control, therefore, I excluded these promoter interactions in this table. Promoter interactions in cluster F are weakened upon CTCF depletion but not entirely lost. Therefore cluster F is shown between parentheses in the decreased column and the maintained column.

	Decreased*	**Similar**	**Increased**
SCC1 depleted	B, C, D, F	A, E, J	K
CTCF depleted	B, C, (F)	A, D, E, J, (F)	G
WAPL RNAi	A, B	C, D, E, F, J	I
Mitosis	A, B, C, D, F	E, J	M

Next, I compared the promoter interactions from the two depleted conditions SCC1 +Auxin and CTCF +Auxin. SCC1 depletion appears to show more promoter interactions with a decreased CHiCAGO score, since interactions in cluster D appear unaffected upon CTCF depletion (see igure 29). Furthermore, promoter interaction scores in cluster F appear more affected by SCC1 depletion than by CTCF depletion.

Subsequently, I evaluated the number of significant promoter interactions per bait between the two depletion conditions. Figure 30 shows the resulting distributions as bar charts. This shows that SCC1 depletion results in a greater loss of promoter

interactions, as well as the number of interactions per promoter. Upon SCC1 depletion, the median number of promoter interactions per bait drops from 4 to 3, whereas this metric shows no change upon CTCF depletion. The mean number of promoter interactions per bait drops from ~7.41 to ~4.61 upon SCC1 depletion, while it drops from ~6.90 to ~6.35 upon CTCF depletion (figure not included). Together, these results show that the used degron system for SCC1 results in a more dramatic loss of promoter interactions than the used CTCF degron system.

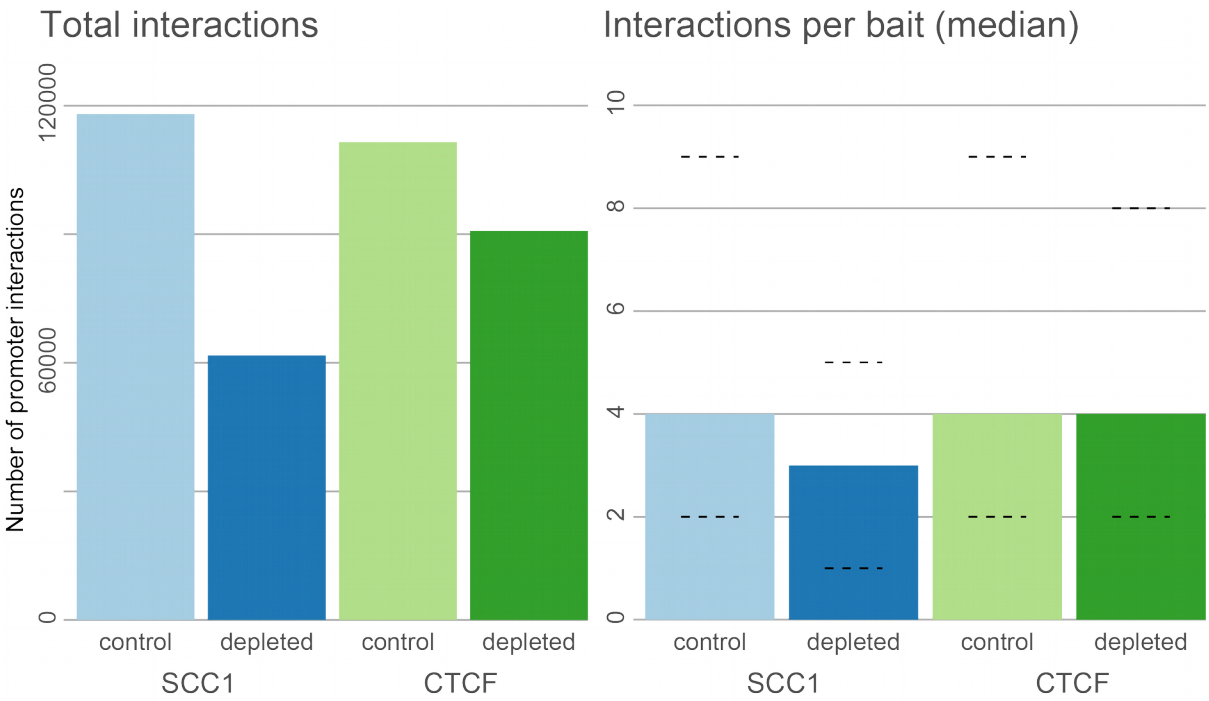

Figure 30. The SCC1-AID system drives greater loss of promoter interactions than the CTCF-AID system. Left: overview of the number of promoter interactions in SCC1 and CTCF control and depleted cells. Right: the median number of promoter interactions per bait in SCC1 and CTCF control and depleted cells (dashed lines indicate Q1 and Q3). Upon SCC1 depletion, the number of promoter interactions decreases by ~47.7% (118,074 to 61,702), whereas upon CTCF depletion, the number of promoter interactions decreases by ~18.6% (111,499 to 90,751). SCC1 depletion shows a greater decrease in the number of promoter interactions per bait than CTCF depletion. SCC1 median number of interactions per bait: 4 (control), 3 (depleted). SCC1 mean number of interactions per bait: ~7.41 (control), ~4.61 (depleted). CTCF median number of interactions per bait: 4 under both conditions. CTCF mean

number of interactions per bait: ~6.90 (control), ~6.35 (depleted). Number of baits per condition: 15925 (SCC1 control), 13397 (SCC1 depleted), 16154 (CTCF control), and 14297 (CTCF depleted). Transchromosomal interactions and bait-to-bait interactions were excluded.

I next sought to detect promoter interaction rewiring using a more formal approach. To this end, in collaboration with William Orchard and Dr Valeriyah Malysheva, we applied the novel differential promoter interaction calling tool Chicdiff (Cairns *et al.*, 2019) that they have developed in the Spivakov lab. Since Chicdiff is currently limited to calling differential promoter interactions between two conditions, two separate runs were performed: SCC1 -Auxin versus SCC1 +Auxin, and CTCF -Auxin versus CTCF + Auxin. I categorized the resulting differential promoter interactions as lost, maintained or gained upon architectural protein depletion (see Methods 2.6). Table 12 and Figure 31 show the total number of interactions per category. These results confirmed the observation from clustering that SCC1 depletion results in greater promoter interaction loss than CTCF depletion, detecting 40666 versus 17645 significantly lost promoter interactions. The numbers of maintained promoter interactions were roughly equal between conditions (14086 versus 13703). Lastly, although gained promoter interactions are relatively rare, they are more than twice as numerous upon SCC1 depletion as upon CTCF depletion (3653 versus 1663).

The numbers of lost, maintained, and gained promoter interactions per bait (Figure 31, right) show that baits commonly lose multiple promoter interactions. This effect is more pronounced upon SCC1 depletion than upon CTCF depletion. Promoters that maintain or gain multiple interactions are comparatively rare. Taken together, these results confirm the extent of previously observed differences in promoter interaction loss and gain. Furthermore, although the similar numbers of maintained promoter interactions between SCC1 and CTCF depleted conditions suggest promoter interaction wiring similarities, a low Jaccard index of ~0.27 indicates that these two sets are actually quite dissimilar (Figure 32). Lastly, these results reconfirm that in the used system, SCC1 depletion leads to more dramatic promoter interaction rewiring than CTCF depletion.

Because SCC1 depletion showed the most dramatic effect in terms of detected promoter interaction rewiring and because cohesin depletion is known to show a distinct reduction in TAD organisation (Haarhuis *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017; Wutz *et al.*, 2017; Nuebler *et al.*, 2018), I focused on cohesin depletion and I largely disregarded CTCF depletion in further analyses.

Table 12. Summary table of the Chicdiff results. Numbers of lost, maintained and gained promoter interactions are shown for two conditions: SCC1 depletion and CTCF depletion. The SCC1 depletion shows more lost, maintained and gained promoter interactions than the CTCF condition.

Condition	Rewiring category	n
SCC1-	Lost	40666
SCC1-	Maintained	14086
SCC1-	Gained	3653
CTCF-	Lost	17645
CTCF-	Maintained	13703
CTCF-	Gained	1663

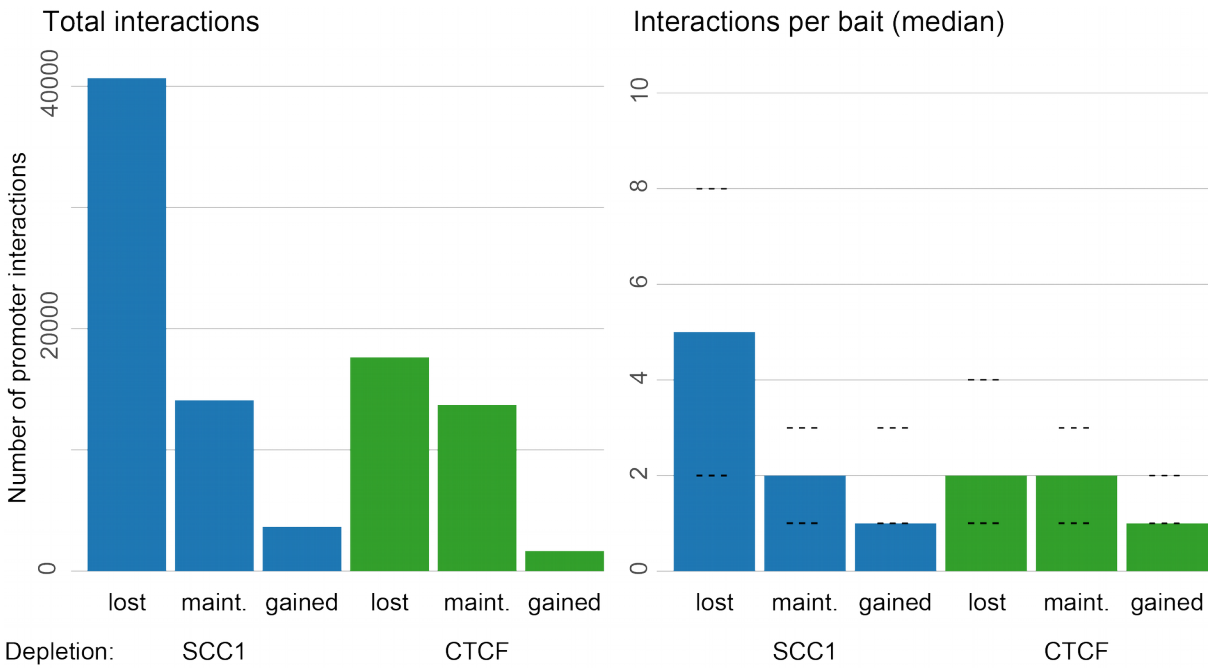

Figure 31. Greater loss of promoter interactions upon SCC1 than upon CTCF depletion. Left: overview of the numbers of lost, maintained and gained promoter interactions upon SCC1 depletion and CTCF depletion. Right: the median number of lost, maintained and gained promoter interactions per bait upon SCC1 and CTCF depletion (dashed lines indicate Q1 and Q3). SCC1 depletion shows more lost promoter interactions than CTCF. The number of maintained promoter interactions is comparable between the two conditions, and the number of gained promoter interactions is slightly higher upon SCC1 depletion than upon CTCF depletion. Upon SCC1 depletion the number of lost interactions per bait is higher than upon CTCF depletion. Transchromosomal interactions and bait-to-bait interactions were excluded.

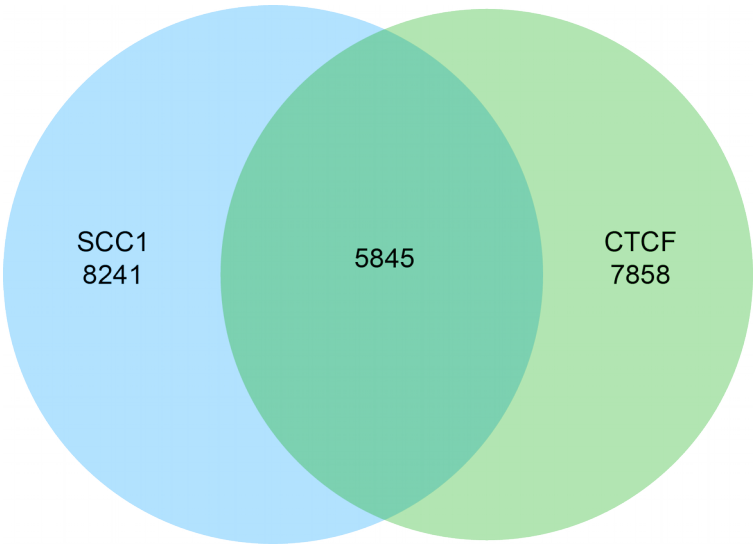

Figure 32. Maintained promoter interactions upon SCC1 and CTCF depletion are largely dissimilar. The number of maintained promoter interactions upon CTCF depletion (left) and SCC1 depletion (right) and their intersection. Although both conditions show similar total numbers of maintained promoter interactions, the intersection between the two sets shows that many different promoter interactions are maintained. This is recapitulated by the Jaccard index value of ~0.27. Transchromosomal interactions and bait-to-bait interactions were excluded.

Since an RF can engage in multiple promoter interactions, I sought to evaluate the extent to which baits and PIRs participate in multiple rewiring categories (e.g. a bait with two promoter interactions; one lost and one maintained). The Venn diagram in Figure 33 shows that baits as well as PIRs commonly exclusively lose, maintain or gain promoter interactions at the same time. However, an appreciably large set of RFs engage in both lost and maintained promoter interactions (~16.8% of baits and ~4.5% of PIRs). Furthermore, although the total number of lost promoter interactions is considerably greater than maintained interactions, the number of baits that engage exclusively in lost or maintained interactions is similar (2153 and 2113 respectively). This suggests that it is more common for promoters to lose multiple interactions than to maintain multiple interactions. Taken together, these results suggest that different interactions of promoters and PIRs may engage in different types of dynamics.

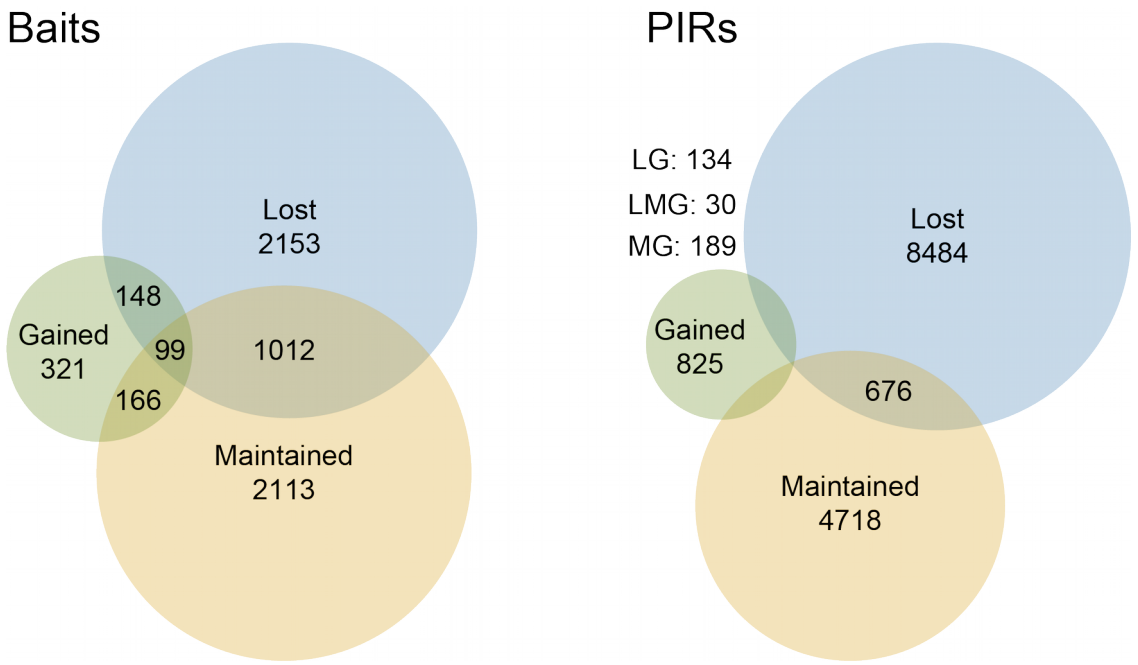

Figure 33. Categorised promoter interactions per bait and per PIR. Baits and PIRs involved in promoter interactions can lose, maintain and gain interactions simultaneously upon SCC1 depletion, however this is a rare occurrence (99 baits and 33 PIRs). It is most common for an RF to respond to SCC1 depletion by exclusively losing or maintaining or gaining interactions. However, a comparatively large set of baits are found to lose as well as maintain promoter interactions (1012 baits). Transchromosomal interactions and bait-to-bait interactions were excluded.

I next aimed to combine the Chicdiff with the clustering results into a consensus set. The rationale for this is as follows. Firstly, K-means clustering does not directly provide a measure of statistical significance of the difference in CHiCAGO scores between conditions. Secondly, although Chicdiff provides a statistical framework for detecting differential interactions, it is limited to analysis between two conditions and is not intended for detecting high-confidence maintained interactions. I reasoned that the K-means approach, which combines data from a multitude of samples, provides valuable information to incorporate into the Chicdiff analysis. I.e., by combining K-means- and Chicdiff-classified promoter interactions, I was able to exclude interactions that show dynamics that are likely unrelated to cohesin depletion (e.g. K-means cluster L in igure 29). I combined K-means- and Chicdiff- classified promoter interactions by the sets of promoter interactions with “decreased”, “similar” and

“increased” CHiCAGO scores upon cohesin depletion (see Table 11), and intersecting these interactions with the Chicdiff sets “lost”, “maintained” and “gained” respectively.

I visualised the proportion of lost, maintained, and gained and non-categorized promoter interactions per cluster (Figure 34). In agreement with the K-means cluster properties, clusters F, J and K most strongly comprise of Chicdiff lost, maintained, and gained promoter interactions respectively. Furthermore, the remaining K-means categorized clusters also largely agree with the Chicdiff categorization and promoter interactions that are unclassified by Chicdiff are revealed to be approximately uniformly distributed among the K-means clusters. I.e. the promoter interactions that are unclassified by Chicdiff closely follow the randomly expected proportion (grey bars in Figure 34). Disagreement between the Chicdiff and K-means classification is also visible: limited numbers of promoter interactions are classified by Chicdiff as lost whereas these interactions are partitioned by K-means into clusters that do not show promoter interaction loss (See yellow bars at A, E, G, H, J, and L, in Figure 34). Additionally, limited numbers of promoter interactions are classified by Chicdiff as maintained whereas these interactions are partitioned by K-means into clusters that do not show promoter interaction maintenance (green bars at F and K in Figure 34). Lastly, limited numbers of promoter interactions are classified by Chicdiff as gained whereas these interactions are partitioned by K-means into clusters that do not show gained promoter interactions (blue bars at A, E, G, J, and M in Figure 34). These discrepancies likely arise because the K-means approach is informed by a multitude of samples. Therefore, the K-means approach classifies a promoter interaction according to its dynamics under a variety of conditions whereas Chicdiff is informed by two conditions only. Therefore, a small but significant change in CHiCAGO score between the SCC1 control and SCC1- conditions may be reported by Chicdiff as promoter interaction rewiring, whereas the additional samples have informed the K-means clustering differently. E.g. Chicdiff classifies promoter interactions in cluster L as lost whereas these interactions appear absent in G1 and G2 and should therefore not be classified as lost. Taken together, the consensus set between K-means and Chicdiff shows broad agreement: although large proportions of the K-means classified promoter interactions did not meet the stringent classification requirements of Chicdiff, the majority of Chicdiff interactions are recovered (lost: 89%, maintained:

92%, gained: 68%, see Figure 35). Furthermore, the promoter interactions that are unclassified by Chicdiff show no bias towards any of the clusters. This likely indicates that these promoter interactions had low signal to noise ratios and subsequently, do not exceed the significance cut off. Lastly, the disagreement between K-means and Chicdiff classification highlights that the K-means clusters inform the consensus classification by discarding promoter interactions based on their rewiring response in additional samples. Henceforth, the consensus rewiring promoter interactions will be referred to as: lost, maintained, and gained.

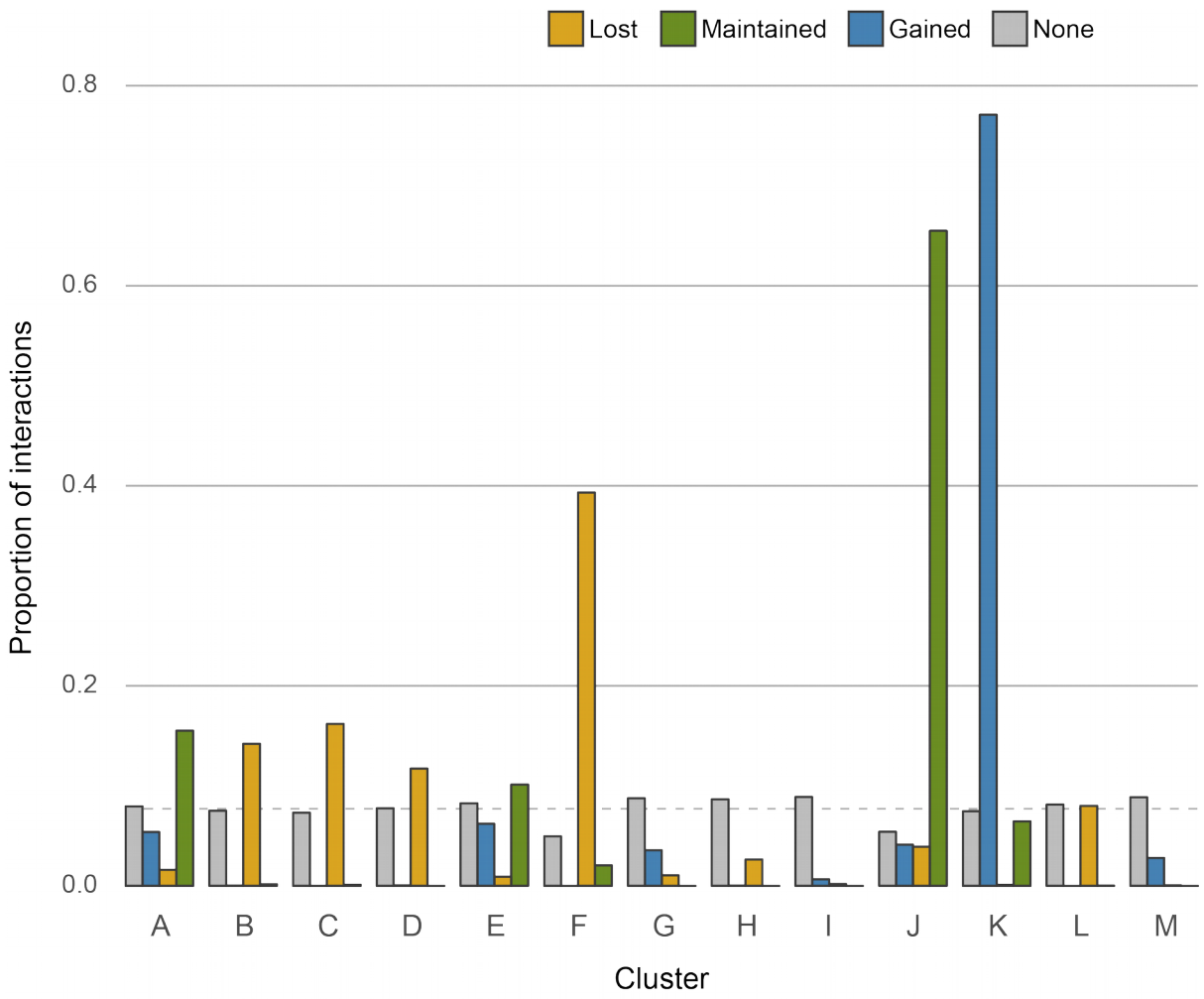

Figure 34. Chicdiff-categorized promoter interactions per K-means class. Lost interactions are overrepresented in clusters B, C, D and F, maintained interactions are overrepresented in clusters A, E and J, and gained interactions are overrepresented in cluster K. Unassigned promoter interactions occur in every cluster at comparable levels. The

dashed line indicates the expected value at random assignment (1/13). Note that the number of interactions of each rewiring category is normalised for cluster size. Transchromosomal interactions and bait-to-bait interactions were excluded.

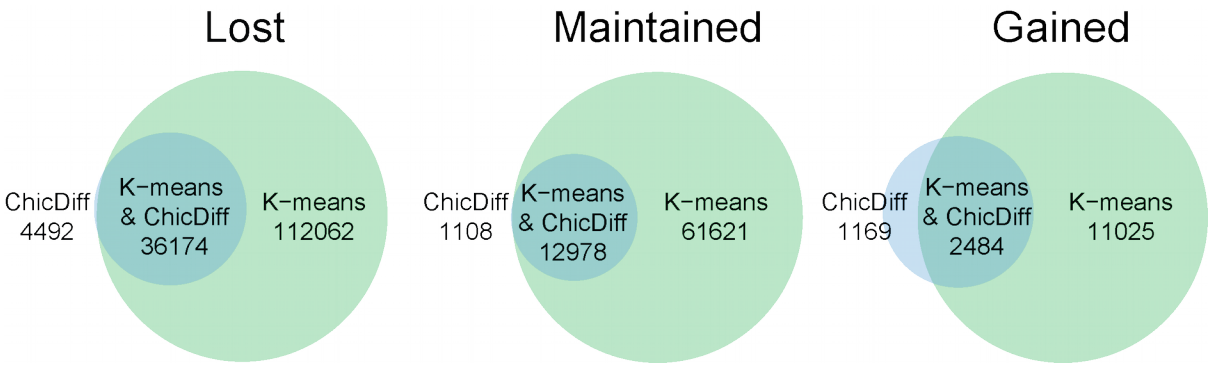

Figure 35. Comparison of K-means and Chicdiff classification. Venn diagrams of Chicdiff and K-means defined lost, maintained and gained promoter interactions between SCC1 control and SCC1 depleted conditions. In all categories, the majority of Chicdiff interactions are identically classified by K-means (Lost: ~89%, maintained: ~92%, gained: ~68%). However, K-means based classification results in larger numbers of interactions than Chicdiff. Transchromosomal interactions and bait-to-bait interactions were excluded.

4.2 Cohesin-dependent promoter interaction rewiring in the context of interphase TAD organisation

Having constructed a consensus set of promoter interactions that rewire upon SCC1 depletion I set out to identify distinct properties of lost, maintained and gained promoter interactions. To this end I first evaluated spatial properties of promoter interactions per rewiring category, and I then compared categorised promoter interactions to TAD organisation.

To determine differences in promoter interaction distance on the linear genome, I visualised promoter interaction distance distributions per rewiring category as violin plots (Figure 36, left). This revealed that lost promoter interactions tend to act over longer genomic distances than maintained and gained promoter interactions. Although maintained promoter interactions tend to act over shorter distances than both lost and gained the distribution of the distances appears bimodal. These results show that there are differences in spatial organisation between the promoter interactions that make up the lost, maintained, and gained categories.

I next sought to ascertain if the observed differences in promoter interaction distance are interaction-specific or inherent to the properties of the interacting promoter. To this end I selected baits that both lose and maintain promoter interactions upon SCC1 depletion. I then visualised the promoter interaction distance as violin plots (Figure 36, right). This recapitulates the interaction distance profiles which show that maintained promoter interactions tend occur over shorter distances even when the bait simultaneously loses longer-range interactions (Figure 36, left). Therefore, I conclude that the differences in promoter interaction distance are inherent to the interactions and not to the promoters.

Since the interaction distance differences do not appear to be driven by promoter-specific properties, I asked what properties of PIRs in cis associate with specific rewiring patterns. I reasoned that active and repressive epigenetic signatures may inform the properties of promoter interactions in the context of cohesin dependent rewiring. To investigate this, I used HeLa specific chromatin annotations from the Ensembl Regulatory Build (Zerbino *et al.*, 2015). These annotations reflect distinct epigenetic signatures and they combine multiple signatures into one annotation: active, poised, repressed or inactive. I integrated these annotations with RFs and selected RFs that contain at least one active annotation or at least one

poised/repressed annotation. I then visualised the promoter interaction distance profiles as violin plots (Figure 37). This shows that RFs involved in maintained promoter interactions that also bear active epigenetic signatures tend to be located significantly farther away from the promoters they contact on the linear genome than RFs of the same interaction category that bear repressed or poised signatures (p = 2.2*10-16). Promoter interaction distances of lost and gained interactions do not appear different when selecting RFs with active versus poised/repressed annotations. This result suggests that promoters with maintained (cohesin-depletion resistant) interactions, contact PIRs with active marks over longer genomic distances than PIRs with poised/repressed marks. I further investigate the properties of PIRs in sections 4.3 and 4.4.

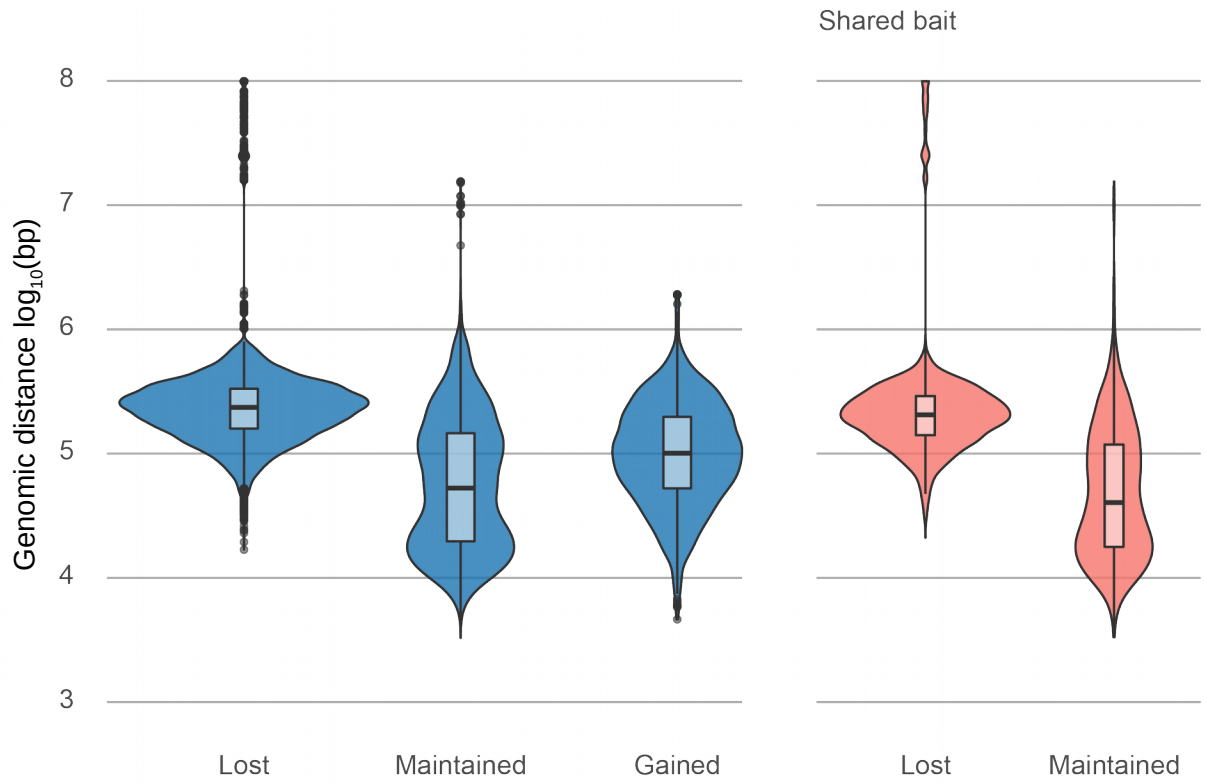

Figure 36. Rewiring promoter interactions show distinct interaction distance profiles. Linear genomic distance of significant promoter interactions stratified by rewiring response upon SCC1 depletion. Left: interaction distance for all significant promoter interactions. Median number of interactions: lost (235,494), maintained (52,770), gained (100,827). Right: distance profiles of baits of promoter interactions that share baits. I.e. in the presence of cohesin, each lost PIR

contacts a bait that is also contacted by a maintained PIR. Number of interactions with shared PIRs: lost (2101), maintained (1787). Transchromosomal interactions and bait-to-bait interactions were excluded.

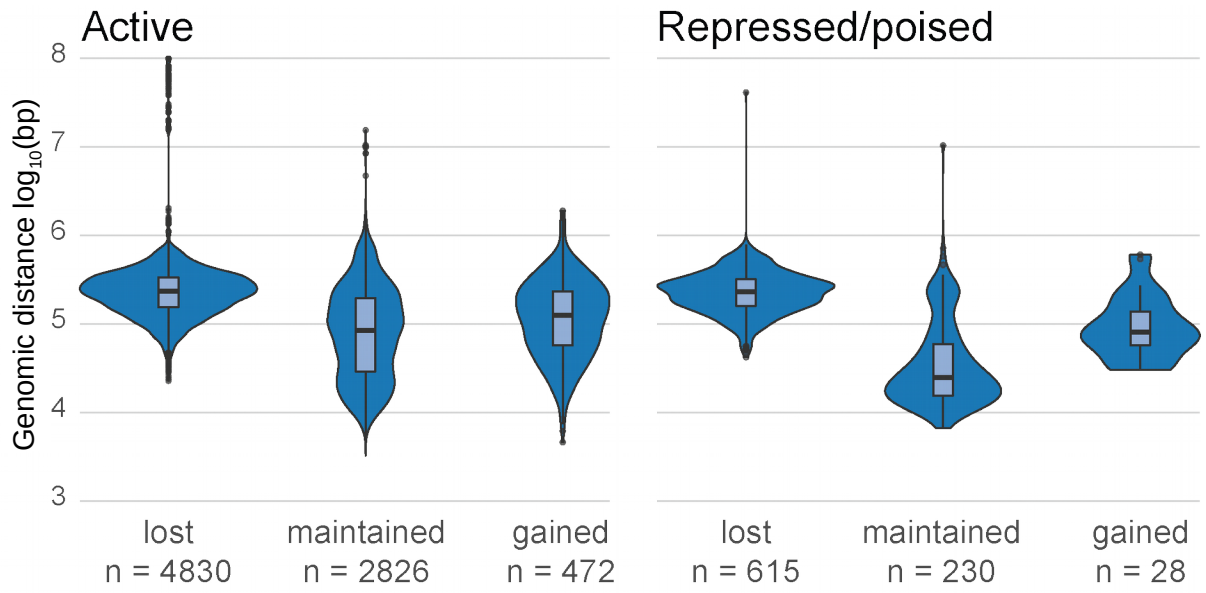

Figure 37. Maintained promoter interactions show distinct interaction distance profiles based on epigenetic annotations. Linear genomic distance (log10(bp)) of significant promoter interactions stratified by rewiring response upon SCC1 depletion. Left: promoter interactions between RFs with at least one active annotation. Right: promoter interactions between RFs with at least one repressed or poised annotation. Epigenetic signature annotations were taken from the Ensembl Regulatory build. A Kolmogorov-Smirnov test on the interaction distances of non-intersecting maintained interactions (active: n = 2797, repressed/poised: n = 201) reported a p-value of 2.2*10-16. Transchromosomal interactions and bait-to-bait interactions were excluded.

I next performed TAD “viewpoint window" analyses (as in the previous chapter) on lost, maintained, and gained promoter interactions. These analyses show the frequency diagrams for three pentile “viewpoint windows”: peripheral, intermediate and central (where the peripheral and intermediate windows combine data from the left as well as the right side of the TAD). Figure 38 shows the viewpoint windows as

columns (from left to right: TAD boundary proximal, intermediate and TAD central) and the rewiring categories as rows (top to bottom: lost, maintained, and gained).

Firstly, the promoter interaction frequency profiles with respect to TAD locations are remarkably distinct per rewiring category. RFs that are involved in lost promoter interactions tend to interact with partner RFs that are located on the opposite side of the TAD. When these RFs are located in the TAD central window, the partner RFs appear at either TAD boundary with similar frequency. In contrast, RFs involved in maintained promoter interactions tend to interact with partner RFs that are located proximally. Both lost and maintained interactions appear constrained by the TAD boundaries. RFs involved in gained promoter interactions tend to interact with partner RFs that are located beyond the most proximal TAD boundary, as expected given that TADs are dissolved upon cohesin depletion.

RFs involved in lost and gained promoter interactions tend to be located proximally to TAD boundaries, which is not the case for RFs involved in maintained promoter interactions. Figure 39 illustrates this observation in terms of enrichment over the uniform distribution of RF locations across each TAD window. This shows that RFs involved in maintained promoter interactions are evenly distributed within TADs. Furthermore, Figure 39 shows that lost promoter interactions are more strongly enriched for PIRs at TAD boundaries whereas gained promoter interactions are more strongly enriched for baits at TAD boundaries. Taken together, these results demonstrate the distinct properties of the lost, maintained, and gained promoter interactions with respect to interphase TAD boundaries. In particular, lost promoter interactions tend to adhere strongly to interphase TAD organisation whereas gained promoter interactions tend to defy it. Furthermore, the RFs involved in maintained promoter interactions can be located anywhere within a TAD, but are generally constrained by TAD boundaries.

Since maintained promoter interactions show different properties when selecting for RFs with active versus poised/repressed annotations, I next performed the TAD viewpoint window analysis exclusively on promoter interactions between RFs with active annotations (Figure 40). The number of RFs in this analysis is comparatively low which results in less clear frequency distributions than the previous analysis. However, the trends are still visible: lost and gained promoter interactions between

active RFs show comparable distributions to the non-selected viewpoint window analysis (in Figure 38). In contrast, the maintained interactions show a different pattern. Namely, the proximal interactions appear absent. This is in accordance with the observation in Figure 37 that RFs with active annotations form longer range promoter interactions than those with poised/repressed annotations. Furthermore, these results indicate that maintained promoter interactions bearing active marks span longer distances than those that do not.

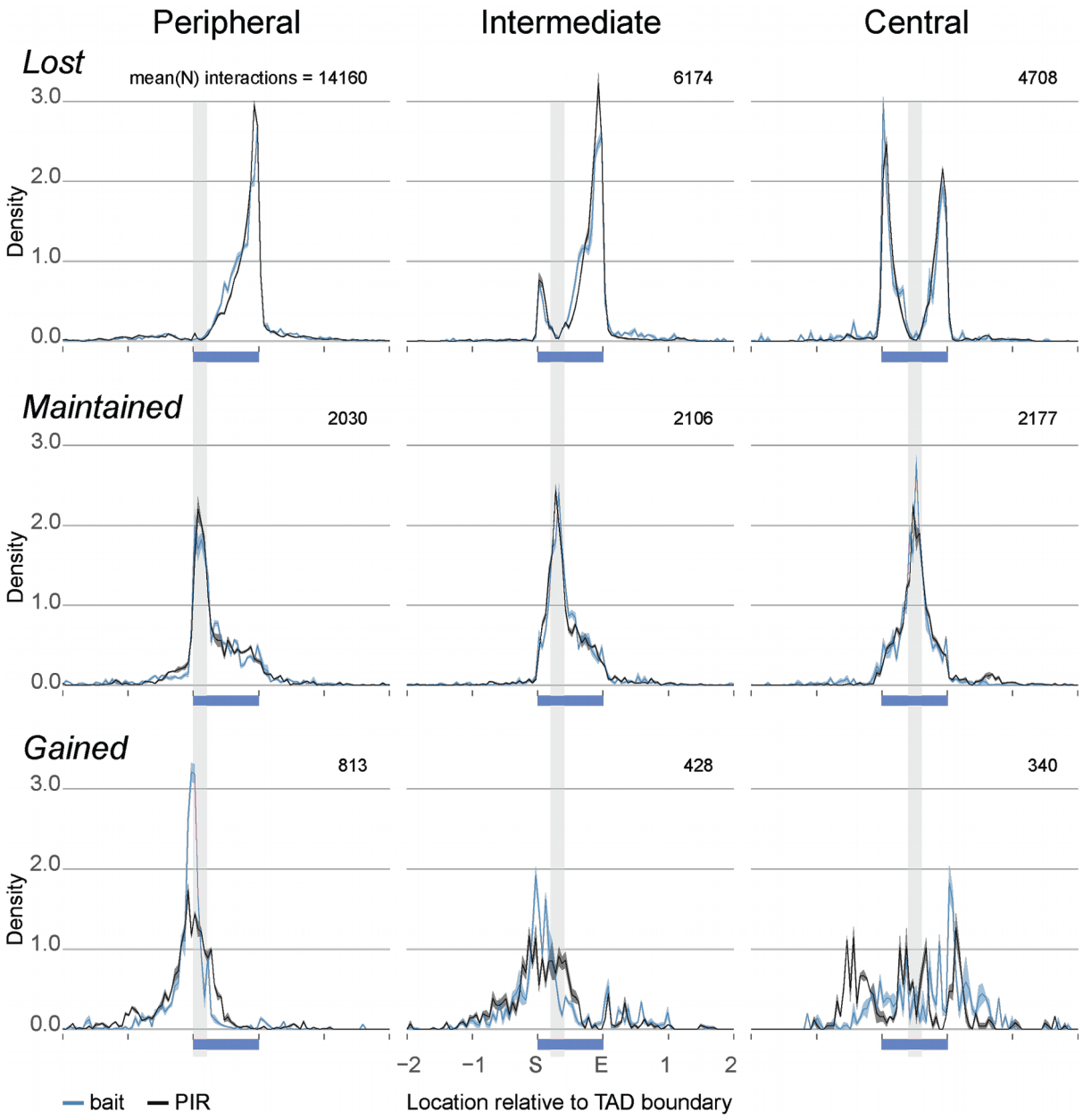

Figure 38. Viewpoint window analysis reveals distinct promoter interaction organisation with respect to TAD organisation. Promoter interactions that are lost, maintained, or gained upon SCC1 depletion are shown on the top, middle and bottom row respectively. The vertical grey bars represent the viewpoint which encompasses TAD boundary proximal (left column), intermediate (middle column), and TAD central (right column) positions. The horizontal blue bars represent TAD intervals. When baits are located within the viewpoint window, the interacting PIRs are represented by the black line. When PIRs are located within the viewpoint window, the interacting baits are represented by the

blue line. The shaded area around the lines represents +/-1SE over four interphase TAD samples. Transchromosomal interactions and bait-to-bait interactions were excluded.

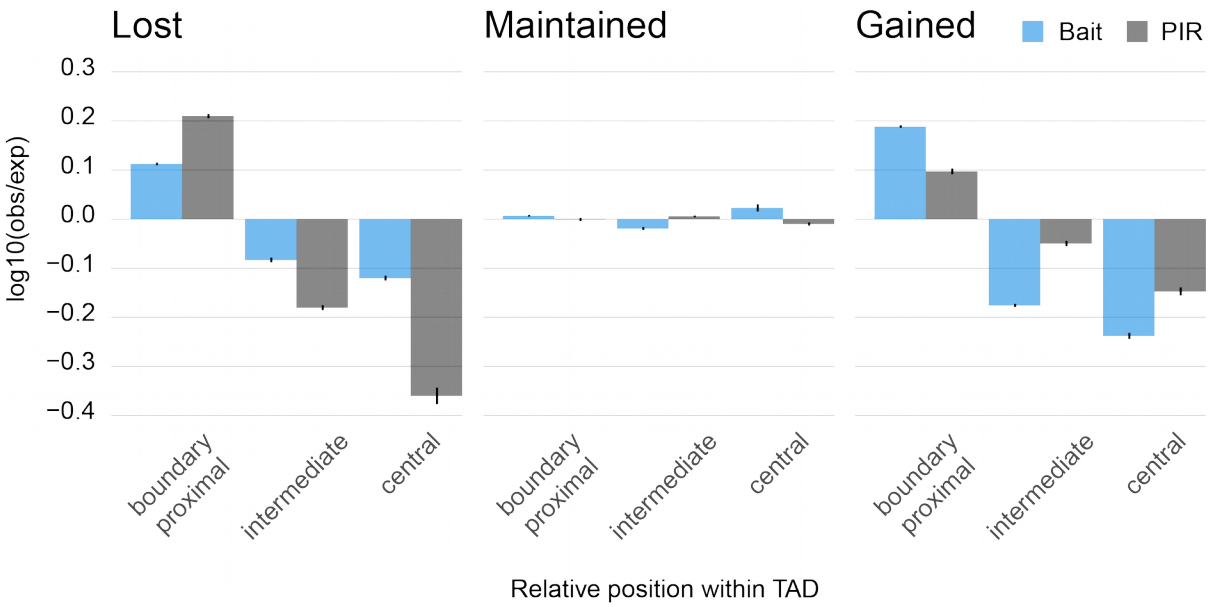

Figure 39. Maintained promoter interactions follow a near-uniform distribution within TADs whereas lost and gained promoter interactions are enriched at TAD boundaries. Enrichment plots showing log10(observed/expected) proportions of RFs within TADs. The uniform distribution is used as the expected distribution. Enrichment is shown separately for baits (blue) and PIRs (grey) involved in lost , maintained, and gained promoter interactions upon SCC1 depletion. Each panel shows three pairs of bars which represent the TAD-boundary-proximal (peripheral), intermediate and TAD-central viewpoints. RFs involved in lost and gained interactions are enriched in the TAD boundary proximal window and depleted in the intermediate and central windows. This property is more pronounced for PIRs than baits of lost interactions, which is not the case for gained interactions. In contrast, RFs involved in maintained promoter interactions show little enrichment, which shows that they are near-uniformly distributed. Note that enrichment is calculated by expecting a uniform distribution. Error bars indicate +/- 1SE over four TAD partitioning replicates. Transchromosomal interactions and bait-to-bait interactions were excluded.

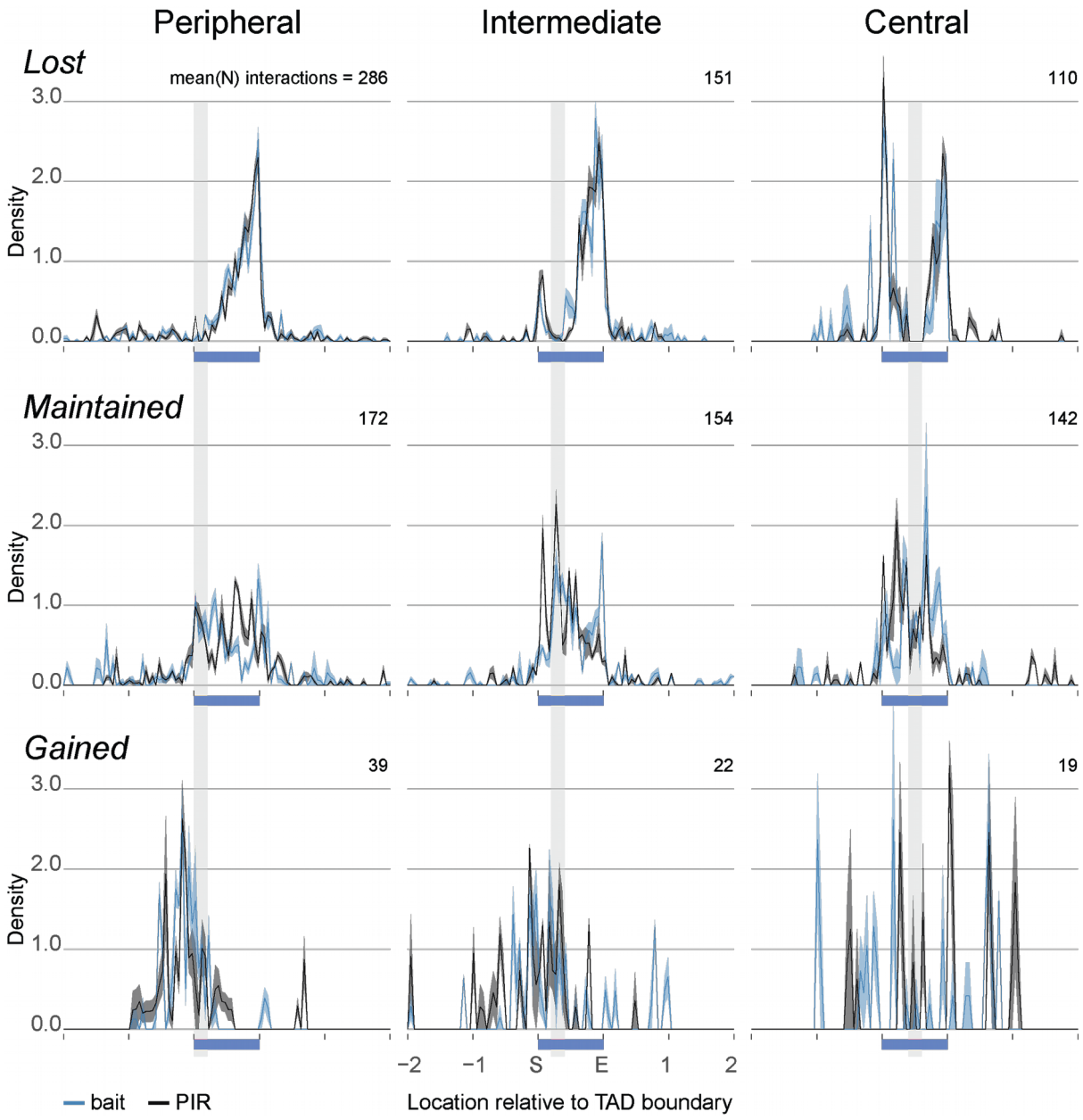

Figure 40. Maintained promoter interactions with active PIRs show fewer short-range interactions. Viewpoint window analysis of "active" promoter interactions that are lost, maintained, or gained (top, middle, bottom row respectively) upon SCC1 depletion. The vertical grey bars represent the viewpoint which encompasses TAD boundary proximal (left column), intermediate (middle column), and TAD central (right column) positions. The horizontal blue bars represent TAD intervals. When baits are located within the viewpoint window, the interacting PIRs are represented by the black line. When PIRs are located within the viewpoint window, the interacting baits are represented by the blue line. The shaded

area around the lines represents +/-1SE over four interphase TAD samples. Note that RFs are only considered when they are unambiguously annotated as “active” in the Ensembl Regulatory build. I.e. the RFs contain at least one region with an active epigenetic signature annotation and none of the remaining annotations: poised, repressed, or inactive. Transchromosomal interactions and bait-to-bait interactions were excluded.

Since promoter interactions appear to show different propensities to cross interphase TAD boundaries per rewiring category, I set out to quantify the TAD-boundary-crossing propensity while controlling for the promoter interaction distance. To this end, I performed a logistic regression analysis where the dependent variable represents TAD boundary crossing (shown in Figure 41). To control for the differential interaction distance patterns between rewiring categories, I included promoter interaction distance as a covariate. This analysis confirmed that newly formed promoter interactions upon SCC1 depletion (*gained*) tend to cross interphase TAD boundaries more frequently than maintained, or lost promoter interactions. Lost promoter interactions display the lowest propensity to cross interphase TAD boundaries, suggesting that this category predominantly comprises intra-TAD promoter interactions. In agreement with previous findings, linear genomic promoter interaction distance is strongly correlated with TAD-boundary crossing. Taken together this analysis reaffirms the observation that gained promoter interactions tend to cross interphase TAD boundaries while adding that lost promoter interactions are unlikely to do so regardless of the fact that these interactions tend to occur over longer genomic distances than maintained or gained promoter interactions.

Since the previous analyses show that the rewiring categories show distinct properties with respect to TAD organisation, I next aimed to find out what proportion of these interactions intersect with detected TADs. My reasoning was that if lost promoter interactions directly reflect TAD organisation and conversely if gained promoter interactions form due to relaxation of TAD boundary insulation, promoter interactions from these two categories must intersect TAD intervals. Furthermore, since maintained promoter interactions appear least affected by TAD dissolution I expected that promoter interactions in this category may exist in regions where no TADs were detected. To investigate this, I calculated the proportion of promoter

interactions per rewiring category that intersect at least one TAD interval. Additionally, I calculated the proportion of the genome over which TAD intervals were detected. I found that the proportion of promoter interactions that intersect TADs is higher for lost and gained than for maintained interactions (~71.9% and ~72.6% versus ~44.8% respectively, Figure 42). The smaller proportion of maintained promoter interactions that intersect TADs indicates that these interactions are frequently located in regions where no high confidence TADs are defined. Figure 42 also shows that the sum of high-confidence TAD intervals covers ~56.5% of the genome. Using standard stringency TADs, this is ~71.7% (not shown in figure). Taken together, these results confirm that lost and gained promoter interactions are more related to high confidence TAD organisation. Additionally, maintained promoter interactions may be less restricted by TAD organisation and may be located in regions of the genome where no high-confidence TADs are detected.

I visualised three examples of differential promoter interactions: *SNORA7* (*lost*), *AL357568.1* (*maintained*) and *LIMCH1* (*gained*). Figure 43Figure 44Figure 45show the respective graphs. These include TAD boundary positions (mean +/- 1SE) of the TAD in which the bait is located. In accordance with the detected frequency distributions (Figure 38), the PIRs of SNORA7 are located proximally to the opposite TAD boundary, the PIRs of AL357568.1 are located proximally to the bait, and the gained PIRs of LIMCH1 are located beyond the interphase TAD boundary.

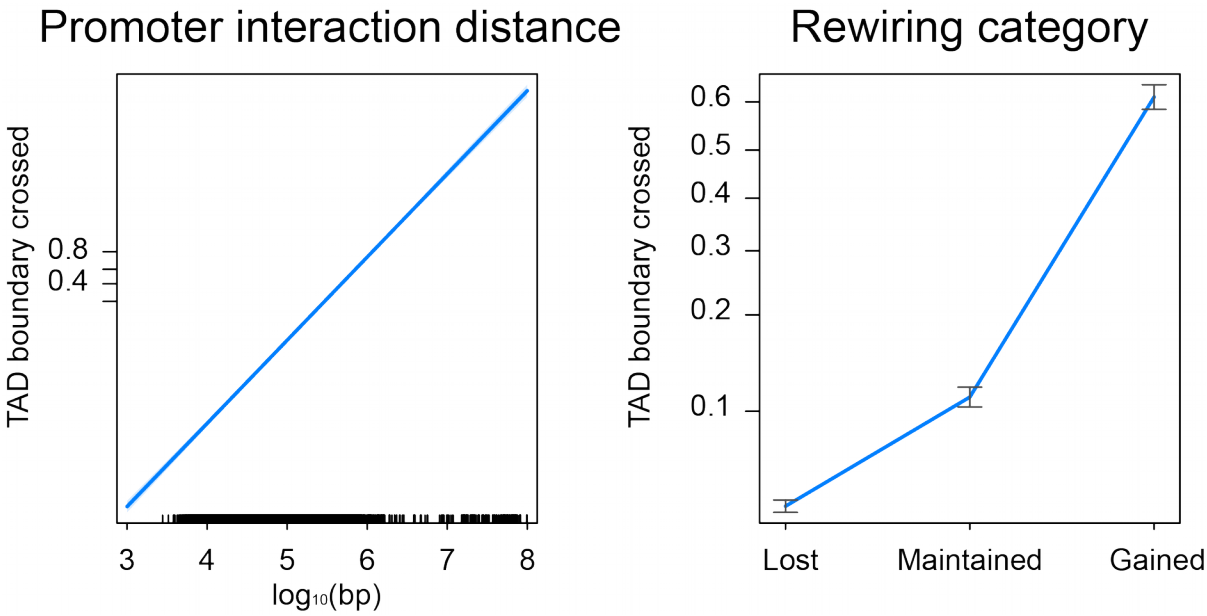

Figure 41. Promoter interaction distance and rewiring category relate to TAD boundary crossing. Effect plot of logistic regression showing TAD boundary crossing versus linear genomic interaction distance and SCC1-dependent promoter interaction rewiring. Left: log10 promoter interaction distance (x-axis) versus TAD boundary crossing (ln odds, y-axis). Right: rewiring category (x-axis) versus TAD boundary crossing (ln odds, y-axis). The odds of a promoter interaction crossing a TAD boundary increase with greater promoter interaction distance. Lost promoter interactions show the lowest odds of crossing a TAD boundary, followed by maintained, and then gained. The independent variables of the regression are the log10 promoter interaction distance and the rewiring category (lost, maintained, gained). The dependent variable is a consensus boolean that is only true if TAD boundary crossing is detected in all four interphase TAD samples. Error bars show a 95% confidence interval.

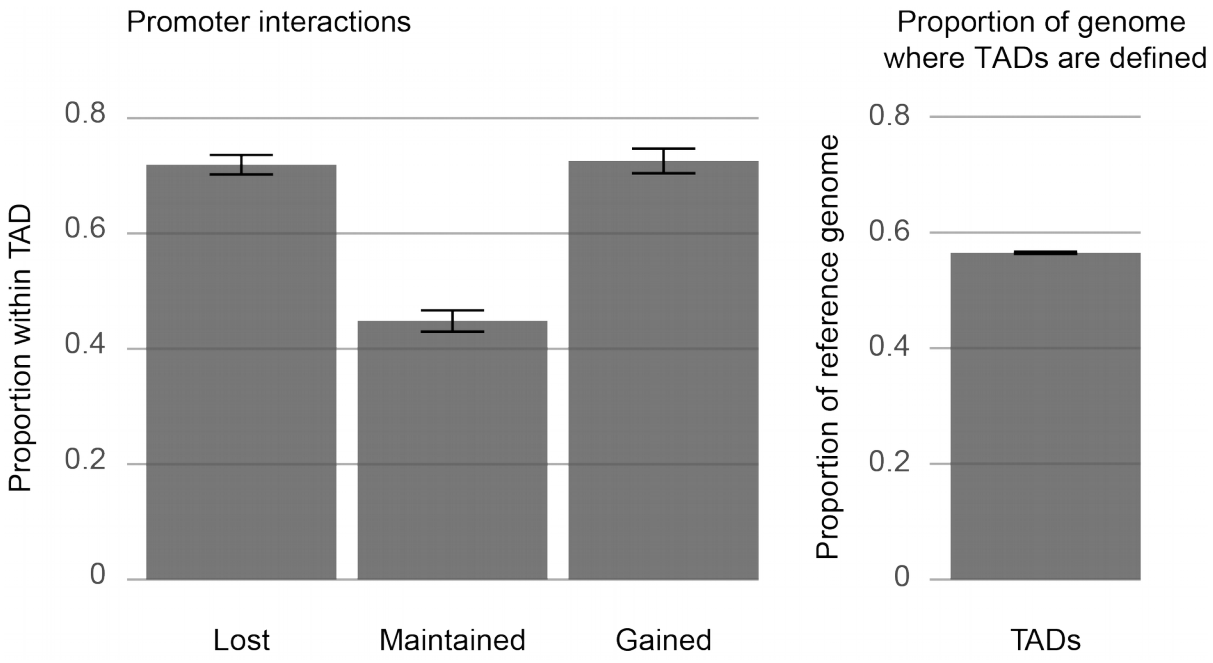

Figure 42. Lost and gained promoter interactions are more related to stringent TADs than maintained promoter interactions. Left: proportion of promoter interactions per rewiring category that intersect at least one TAD. ~71.9% of the lost and ~72.6% of the gained promoter interactions intersect TADs whereas this is ~44.8% for maintained promoter interactions. TAD intersection is defined as the proportion of promoter interactions with the bait or the PIR intersecting at least 1 bp with an interphase TAD interval. Right: high-confidence TAD intervals are defined over ~56.5% of the reference genome (GRCh37). Error bars show +/- 1SE between 4 interphase TAD samples.

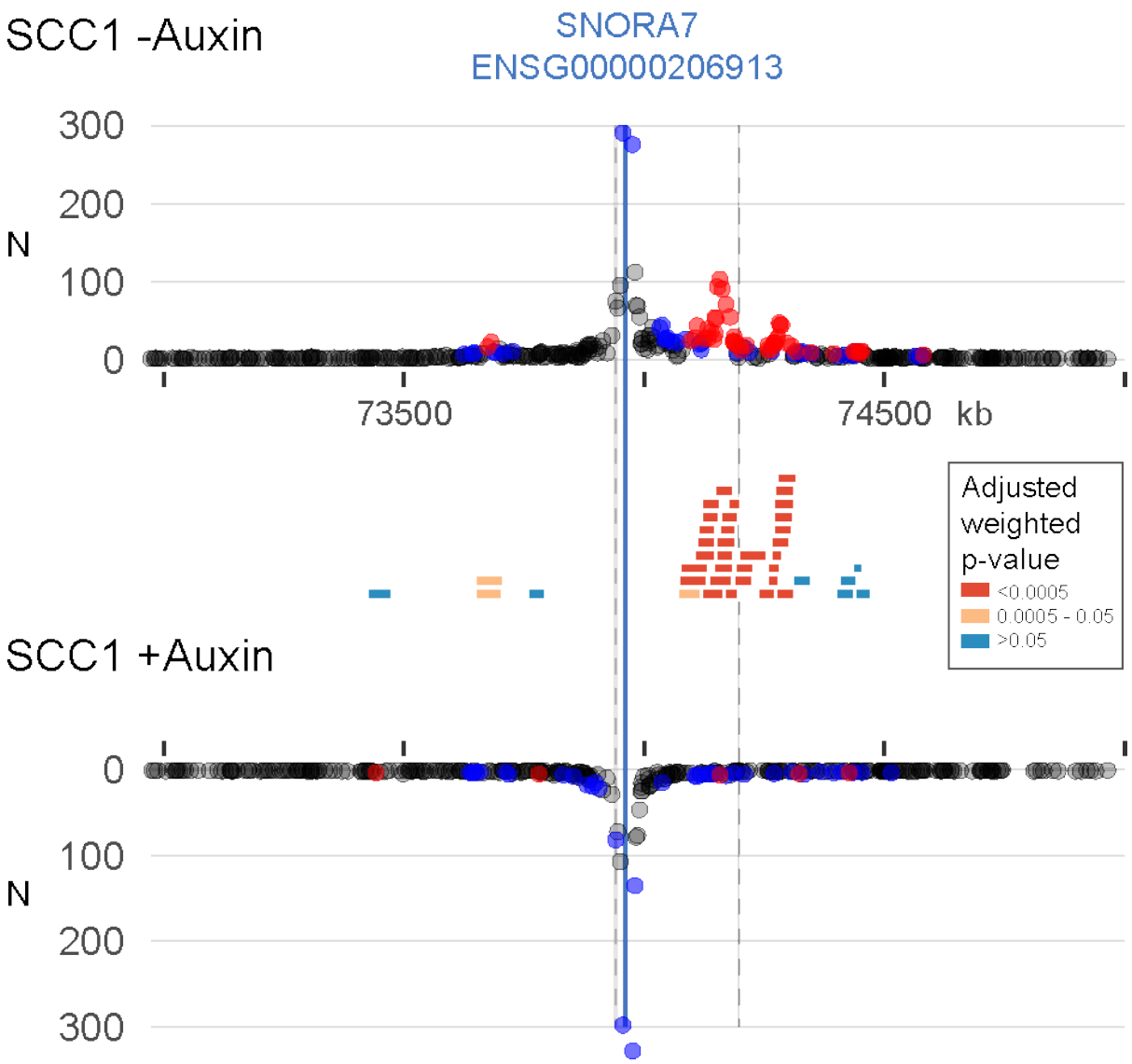

Figure 43. Example of lost promoter interactions upon SCC1 depletion. The boundaries of the most proximal TAD are shown as dotted grey lines with a shaded area that represents +/-1SE over 4 interphase TAD samples. Under -Auxin control conditions (top), the bait containing the promoter of SNORA7 (blue line) shows numerous significant interactions (red circles, CHiCAGO score >= 5). The PIRs with the highest amplitude are located proximally to the opposite TAD boundary of the bait. Under +Auxin SCC1 depleted conditions, the amplitude of the PIRs decreases. Coloured bars represent the Chicdiff adjusted, weighted p-values (see legend), showing that the lost promoter interactions are highly significant.

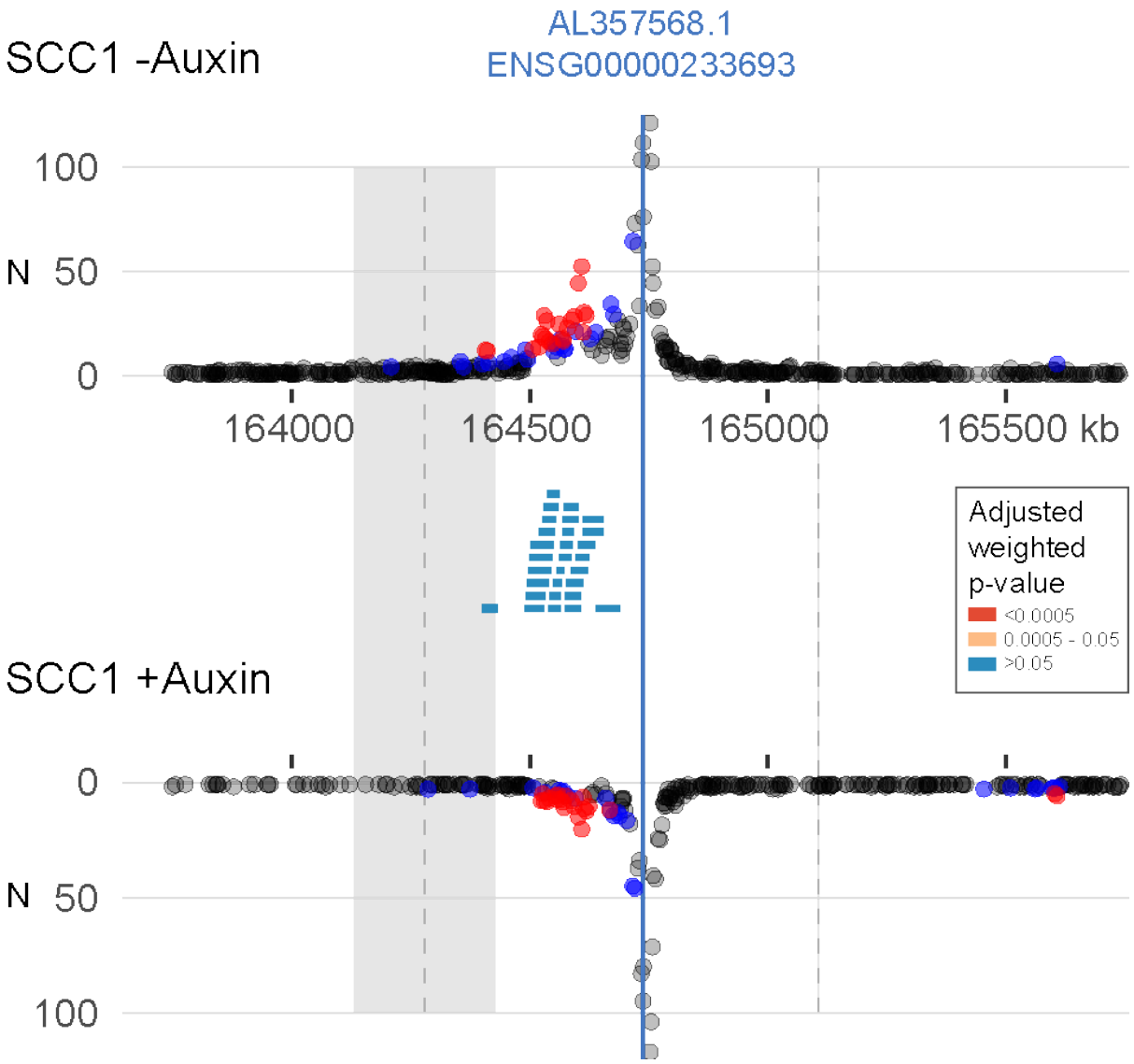

Figure 44. Example of maintained promoter interactions upon SCC1 depletion. The boundaries of the most proximal TAD are shown as dotted grey lines with a shaded area that represents +/-1SE over 4 interphase TAD samples. Under both conditions, the bait containing the promoter of AL357568.1 (blue line) shows numerous significant interactions (red circles, CHiCAGO score >= 5). The PIRs with the highest amplitude are located centrally in the TAD and proximal to the bait. Under +Auxin SCC1 depleted conditions, the amplitude of the PIRs decreases marginally. Coloured bars represent the Chicdiff adjusted, weighted p-values (see legend), showing that the maintained promoter interactions show no significant difference between Auxin+ and Auxin- conditions.

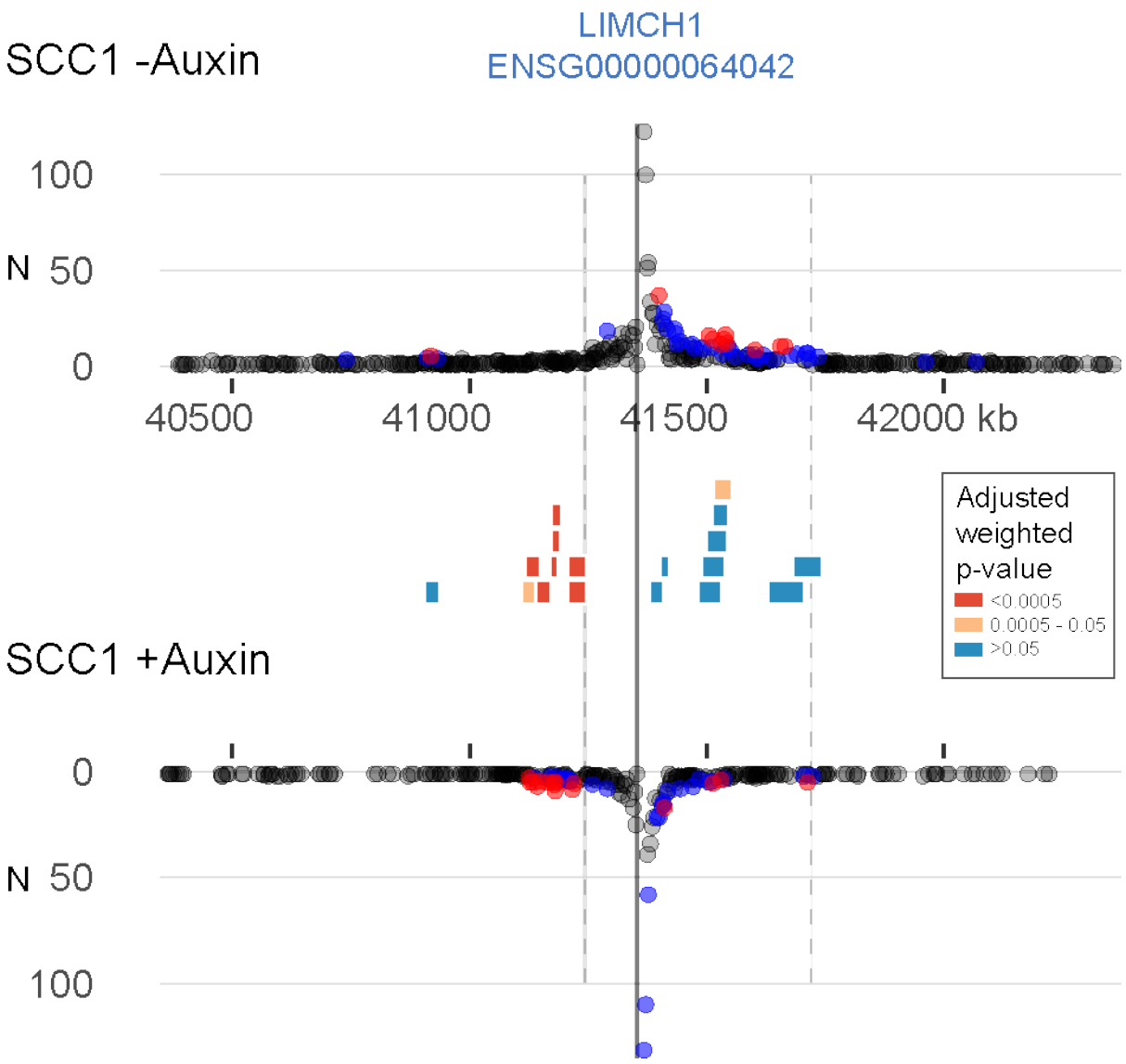

Figure 45. Example of gained promoter interactions upon SCC1 depletion. The boundaries of the most proximal TAD are shown as dotted grey lines with a shaded area that represents +/-1SE over 4 interphase TAD samples. Under both conditions, the bait containing the promoter of LIMCH1 (grey line) shows significant interactions (red circles, CHiCAGO score >= 5). The PIRs with the highest amplitude are located within TAD boundaries when SCC1 is present (top) but are located beyond the left TAD boundary under SCC1 depleted conditions (bottom). Coloured bars represent the Chicdiff adjusted, weighted p-values (see legend), showing that the gained promoter interactions are highly significant. Additionally, some minor interaction loss/maintenance can be seen to the right of the bait (roughly at the 41500kb mark).

4.3 Features of promoter interactions in the context of rewiring upon cohesin depletion

In the previous section I focused on the rewiring of promoter interactions upon cohesin depletion. In this section I explore DNA binding and histone modifications at RFs involved in the rewiring promoter interactions. To this end I use the in-house developed approach of integrating ChIP-Seq data at the RF level. This approach is not based on discrete binding peaks detected with conventional ChIP-seq peak-calling software, but rather, integrates the ChIP-seq signal over entire RFs and expresses this in the form of statistically meaningful scores (see Methods for details). I developed this method specifically for use with restriction-fragment-based analyses such as PCHiC and Hi-C. See Supplementary Figure S 1 – S 4 for a comparison between MACS2 and the per-RF approach used here.

I first tested whether cohesin and CTCF binding are relevant predictors of promoter interaction rewiring upon cohesin depletion by performing two logistic regression analyses. The dependent variable in both analyses represents the classification of an RF as maintained or lost. The independent variables are the log10 promoter interaction distance, the target ChIP-Seq score at the bait and the target ChIP-Seq score at the PIR. The interaction effect between the last two variables is also included in the model. Note that the promoter interaction distance was included to account for distance related effects. The resulting regression (Figure 46) shows that cohesin and CTCF binding at baits are significant negative predictors of promoter interaction maintenance. I.e. increased cohesin or CTCF presence at a bait relates to a decreased likelihood that a promoter interaction is maintained upon cohesin depletion and an increased likelihood that the promoter interaction is lost. This result indicates that lost promoter interactions are preferentially supported by cohesin/CTCF binding in cis.

I then asked whether cohesin and CTCF bind directly to the RFs involved in rewiring promoter interactions. Since lost promoter interactions appear closely related to TAD organisation, I expect that the corresponding RFs show higher levels of cohesin and CTCF binding. Figure 47 shows cumulative density plots of the ChIP-Seq scores for cohesin (left) and CTCF (right). This reveals that RFs involved in maintained promoter interactions tend to show lower scores for both cohesin and CTCF. RFs involved in lost and gained promoter interactions show very similar cumulative

density curves. Overall, the differences between the three rewiring categories are limited. This indicates that cohesin and CTCF are not exclusively bound at RFs involved in lost promoter interactions. This is unsurprising because cohesin and to a lesser degree CTCF can be found ubiquitously throughout the genome, and because I found that RFs can be involved in lost, maintained and gained promoter interactions simultaneously (See Venn diagram in Figure 33). Collectively, this shows that RFs involved in lost and gained promoter interactions tend to have high CTCF/cohesin scores more frequently than RFs involved in maintained interactions.

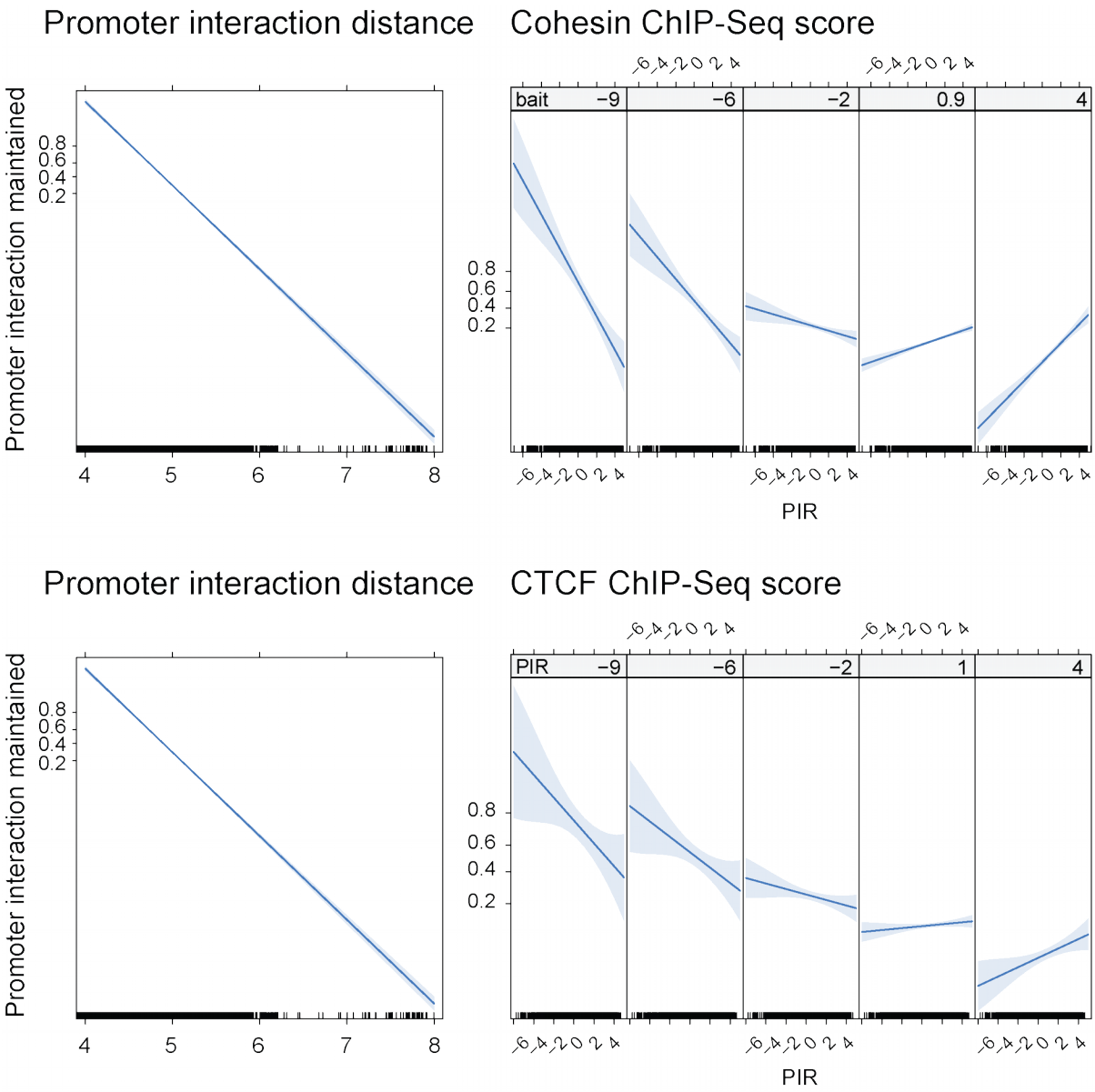

Figure 46. Maintained promoter interactions tend to have reduced cohesin and CTCF ChIP-Seq signal at the bait as well as the PIR. Logistic regression effects plots relating promoter interaction maintenance upon cohesin depletion to interaction distance (between the bait and PIR) and cohesin (top) or CTCF (bottom) ChIP-Seq signal. The dependent variable represents promoter interaction maintenance. I.e. it is true when a promoter interaction is maintained but it is false if an interaction is lost. The independent variables are the promoter interaction distance log10|bp| and the ChIP-Seq signal for cohesin (top) and CTCF (bottom). Note that the inclusion of promoter interaction distance as an independent variable controls for the distance properties of the rewiring

categories. The cohesin score at baits and PIRs are both significant predictors of maintained promoter interactions ($p = 2*10\text{-}16$, $p = 6.4*10\text{-}3$ respectively), though the sign of the beta coefficients differs ($\beta \approx -0.299$ and $\beta \approx 0.065$ respectively). The interaction term is also significant ($p = 3.2*10\text{-}12$) and the beta coefficient is slightly positive ($\beta \approx 0.099$). The CTCF score at baits (but not at PIRs) is a significant predictor of maintained promoter interactions ($p = 2*10\text{-}16$). The interaction term is also significant ($p = 1.52*10\text{-}2$) and the beta coefficient is slightly positive ($\beta \approx 0.038$).

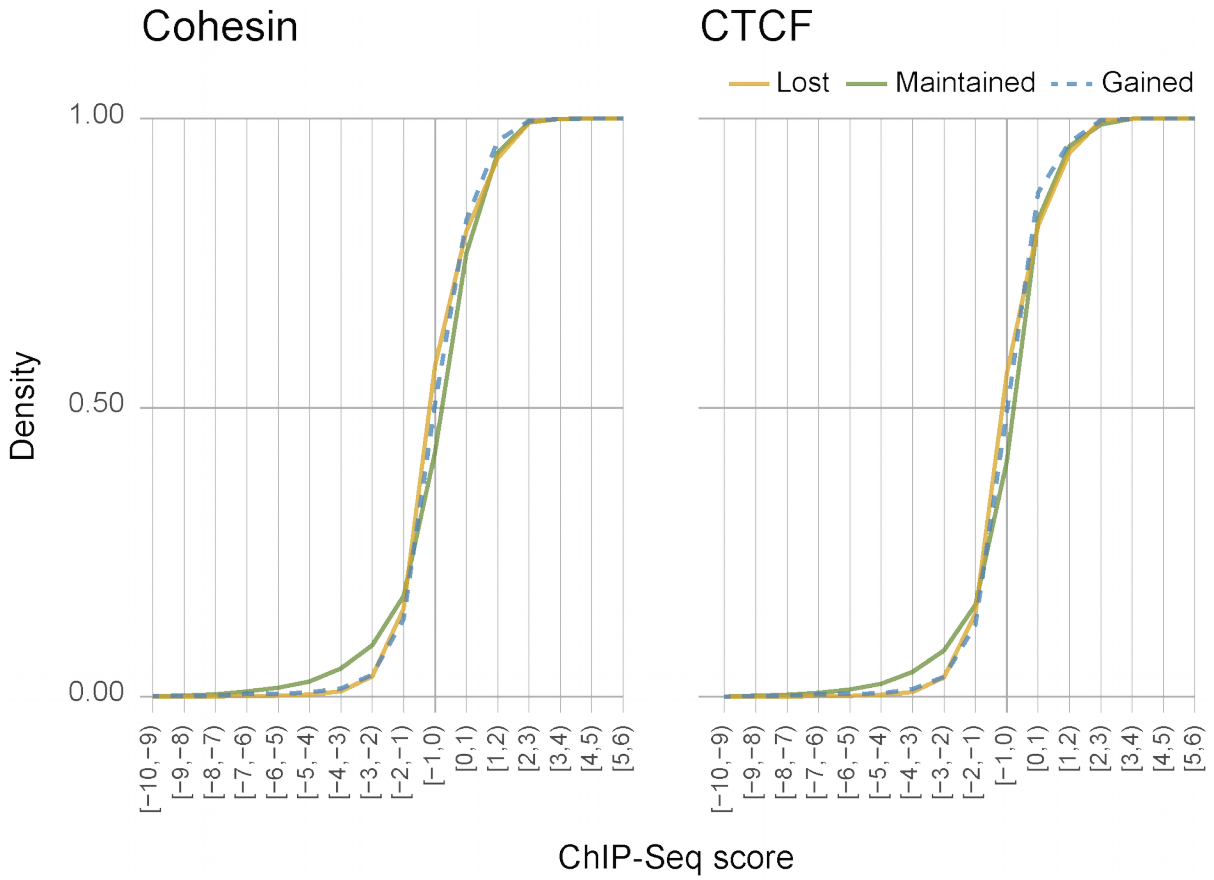

Figure 47. Maintained promoter interactions show a slight decrease in Cohesin and CTCF ChIP-Seq signal at interacting restriction fragments. Cumulative density plots of cohesin (left) and CTCF (right) ChIP-Seq signal at RFs involved in lost, maintained, and gained promoter interactions (baits as well as PIRs). The x-axis shows binned ChIP-Seq residual scores that were calculated per RF. The y-axis shows the cumulative density of the ChIP-Seq scores.

I next sought to identify additional DNA binding proteins that correlate with lost or maintained promoter interactions. I reasoned that multivariate regression may not be the most suitable approach here because I aim to analyse a multitude of ChIP-Seq targets while being chiefly interested in those that show strong effects. Additionally, the ChIP-Seq signals are likely correlated. Therefore, I opted to use logistic regression with LASSO which solves these two issues by shrinking regression coefficients towards zero by means of the penalty factor lambda.

I compiled a compendium of HeLa ChIP-Seq data from publicly available resources as well as from in-house data from my collaborators, which after rigorous quality control resulted in ChIP-Seq data for 51 target proteins (see Methods 2.10). I then performed LASSO logistic regression analyses where the dependent variable represents the rewiring response upon cohesin depletion (maintained versus lost). Gained interactions were not considered in this analysis. The independent variables were the ChIP-Seq scores per PIR in addition to the log promoter interaction distance and the processed promoter interaction score. Figure 48 shows the resulting regularization paths, which reveal 9 target proteins to significantly associate with PIRs of maintained promoter interactions and 7 target proteins to significantly associate with PIRs of lost promoter interactions. Consistent with the conventional logistic regression analysis presented above, promoter interaction distance, cohesin and CTCF are found to correlate with PIRs that are involved in lost promoter interactions. Conversely, the promoter interaction score correlates with maintained interactions. This indicates that strong promoter interactions are more likely to resist the spatial distortions that accompany cohesin depletion, suggesting that an independent mechanism supports these promoter interactions.

The Gene Ontology analysis tool LAGO revealed that the factors that associate with PIRs of maintained promoter interactions are involved in transcriptional regulation[1]. All of the targets were annotated with the GO term “regulation of transcription, DNA-templated” or a related term. Other notable terms were:

- “Chromatin organization”: all factors except TBP
- “Transcription by polymerase II”: all factors except PHF8

[1] LAGO found no annotation for PWWP2A so this is not included in the presented results.

- “Positive regulation of gene expression”: P300, PHF8, and TBP
- “Negative regulation of gene expression”: EZH2, MBD3, and PHF8

Taken together, these GO annotations indicate that PIRs involved in maintained promoter interactions associate with RNA polymerase 2 related transcription factors. This in turn suggests that these PIRs contain enhancers and, importantly, that transcriptional machinery may be supporting the promoter interaction in the absence of cohesin. Lastly, it should be noted that I found activating as well as silencing proteins at PIRs involved in maintained promoter interactions. This is in accordance with the previous identification of distinct subsets of PIRs with active versus poised/repressed Regulatory Build annotations (Figure 37). See discussion 5.2.2 for more detail on protein binding at PIRs in the context of promoter interaction rewiring.

LAGO Gene Ontology analysis of the factors that associate with PIRs of lost promoter interactions yielded unclear results: all seven factors are annotated with the GO term “nucleobase-containing compound biosynthetic process” or a related term. This high-level GO term indicates that these factors are involved in the synthesis of nucleobases, nucleosides, nucleotides or nucleic acids. However, without a more specific annotation, this is not very informative. Other notable terms were:

- “Chromosome organization”: CTCF, HCFC1, RPA1, RPA2, SCC1
- “Protein localization to chromosome”: CTCF, RPA1, RPA2, SCC1
- “Protein-DNA complex subunit organization”: BRF2, CTCF, RPA1, RPA2

The presence of the term “chromosome organization” likely reflects the known involvement of cohesin and CTCF in TAD organization and it suggests that HCFC1, RPA1 and RPA2 are present at TAD boundaries. HCFC1 has been previously shown to associate strongly with cohesin and CTCF (Whalen *et al.*, 2016). However, the DNA repair proteins RPA1 and RPA2 have not.

To identify which histone proteins and histone modifications associate with PIRs involved in maintained versus lost promoter interactions, I performed an additional LASSO logistic regression analysis. Figure 49 shows the resulting regularization paths which reveal that PIRs involved in maintained promoter interactions associated with H3K27ac, H3K4me1, and H3K4me3. The first two of these modifications are typically found at enhancers (Bernstein *et al.*, 2005; Heintzman *et al.*, 2009; Shlyueva *et al.*, 2014). LAGO analysis of H3ac and H2AFZ found no enrichment for

any GO term. Taken together, the histone marks at PIRs involved in maintained promoter interactions corroborate the previous observation that maintained PIRs appear to contain enhancers (see Discussion 5.2.2).

To further corroborate the findings from the LASSO logistic regression analyses, I sought to analyse which rewiring category is overrepresented for RFs with strong ChIP-Seq signal (≥95th percentile) for each of the targets that I found with the LASSO regression. To this end, I focused on the targets that I detected as significant predictors in the LASSO analyses, in addition to the Pol II subunit POLR2A[2]. I constructed a contingency table for each target. These contain the tallies of RFs per rewiring category that exceed the 95th percentile of ChIP-Seq signal versus those that do not (i.e. two by three contingency tables with rows: ≥95th percentile, <95th percentile; and columns: lost, maintained, gained). I subsequently calculated log-odds ratios from these contingency tables. These show to what extent each rewiring category is overrepresented in terms of strong ChIP-Seq signal. I performed Fisher exact tests for each contingency table (i.e. each LASSO-identified ChIP-Seq target) individually, followed by Benjamini-Hochberg correction. I then constructed log-odds ratio heat maps, including only significant targets (FDR ≤ 0.05). Figure 50 shows the resulting heat maps for baits (left) and PIRs (right), which reveals that 8/9 maintained-associated protein targets and all maintained-associated histone targets from the LASSO regression were identically classified by the 95th percentile analysis. The results on lost-associated protein and histone targets do not agree as well with the lasso regression: 2/7 protein targets and 0/2 histone targets were similarly classified (see Discussion 5.2.2). In accordance with my previous findings, I found cohesin and CTCF to be overrepresented at baits and PIRs of lost promoter interactions. However, I also found EZH2 at these baits and PIRs. This is surprising since the lasso regression shows EZH2 to be present at PIRs of maintained promoter interactions. A possible explanation for this discrepancy lies in the differences of the two analysis techniques. Since LASSO regression eliminates highly collinear variables by design, it provides a sparse model where the influence of promoter interaction distance and CHiCAGO score are controlled for by including these as independent variables. Therefore, the LASSO analyses provide more controlled and likely more robust results. In the 95th percentile analysis, I take a

[2] I included POLR2A because transcriptional activity was a common property of the targets detected by lasso regression.

thresholded approach combined with a Fisher exact test. This does not account for promoter interaction distance, CHiCAGO score, or predictor multicollinearity. It should be noted that the 95th percentile analysis aims to identify strong binding rather than any binding. E.g. the observed discrepancies may indicate that EZH2 is present at PIRs of lost as well as maintained promoter interactions; strongly binding at PIRs of lost interactions while exhibiting weaker binding at PIRs of maintained interactions. Furthermore, the 95th percentile analysis provides confirmation of most factors involved in maintained promoter interactions while suggesting that RFs involved in maintained and gained promoter interactions appear similar in terms of overrepresentation for RFs that strongly bind the selected ChIP-Seq targets (see Figure 50).

Baits and PIRs show some dissimilarities in the 95th percentile analysis: four targets are significantly overrepresented at PIRs but not at baits: H3K4me1, RPA1, RPA2 and BRF2. The first substantiates that maintained PIRs exhibit enhancer like characteristics, since H3K4me1 is an enhancer-specific mark. The latter three cluster together with GABPA and HCFC1 and appear more strongly overrepresented at RFs involved in gained than maintained promoter interactions.

Taken together, the 95th percentile analysis largely reflects the identified targets at RFs of maintained promoter interactions. Furthermore, this analysis shows that maintained and gained promoter interactions show similarities in terms of strong ChIP-Seq binding for the selected targets. Since many of these ChIP-Seq targets relate to transcriptional regulation (see LAGO results above), in addition to the previous observation that PIRs of maintained promoter interactions appear to display enhancer properties, this suggests that PIRs of gained promoter interactions are similarly enhancer-like.

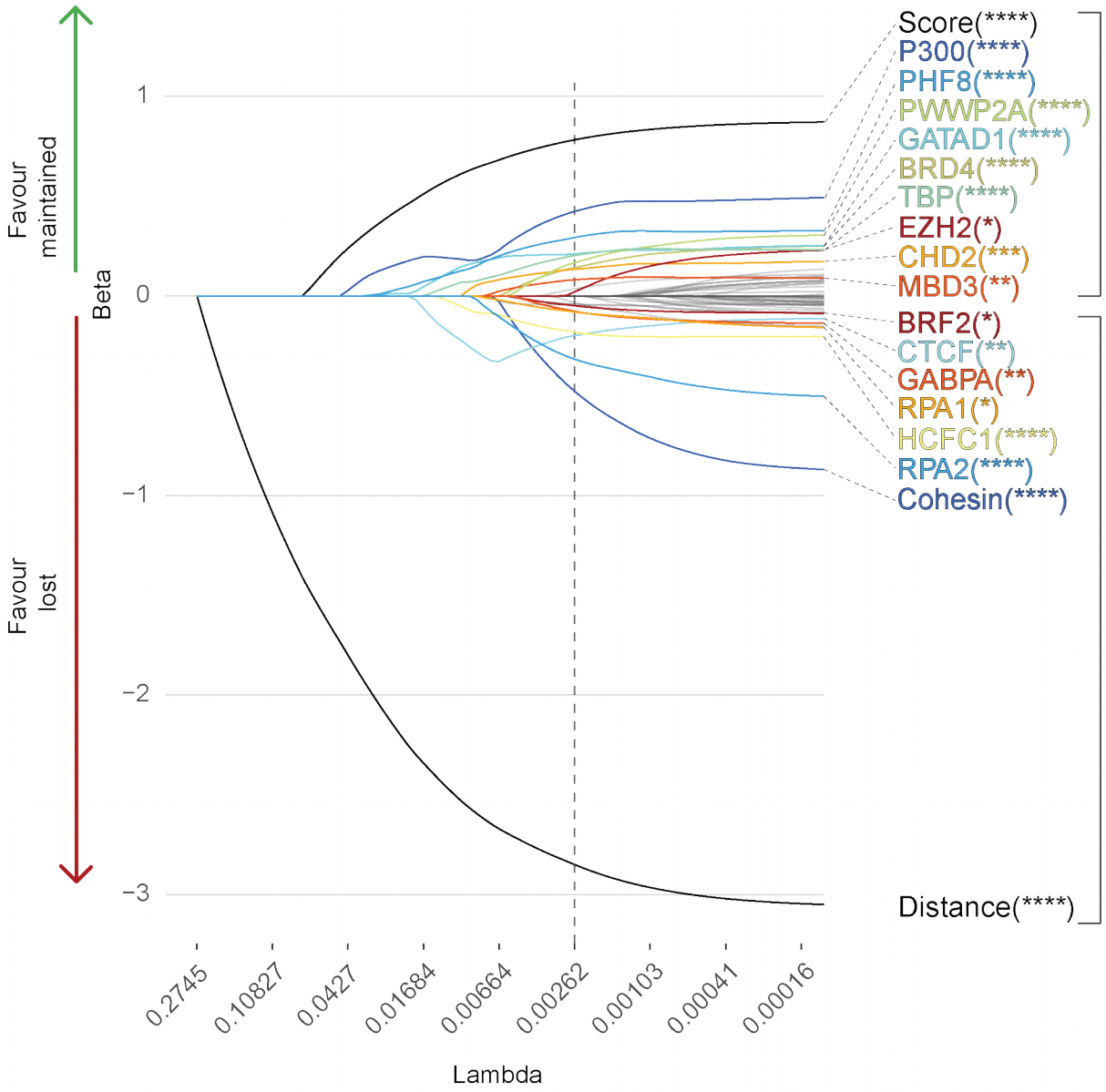

Figure 48. Lasso logistic regression regularization path of DNA binding proteins versus promoter interaction rewiring upon SCC1 depletion. The dependent binary variable represents maintained or gained promoter interactions. The independent variables comprise the ChIP-Seq scores for 51 DNA binding proteins, in addition to the CHiCAGO score and the linear genomic promoter interaction distance ($\log_{10}$(bp)). The y-axis shows the regression beta which is positive for maintained promoter interactions and negative for lost promoter interactions. Note that gained promoter interactions are not taken into account in this analysis. The x-axis shows the lasso penalty coefficient, ranging from stringent on the left to permissive on the right. The vertical dotted line

shows the lambda value at which the independent variables are tested for significance. 9 proteins are found to relate to promoter interaction maintenance, 7 to promoter interaction loss (including cohesin and CTCF). Asterisks indicate significance: **** $p < 0.0001$, *** $0.0001 < p \leq 0.001$, ** $0.001 < p \leq 0.01$, * $0.01 < p \leq 0.05$.

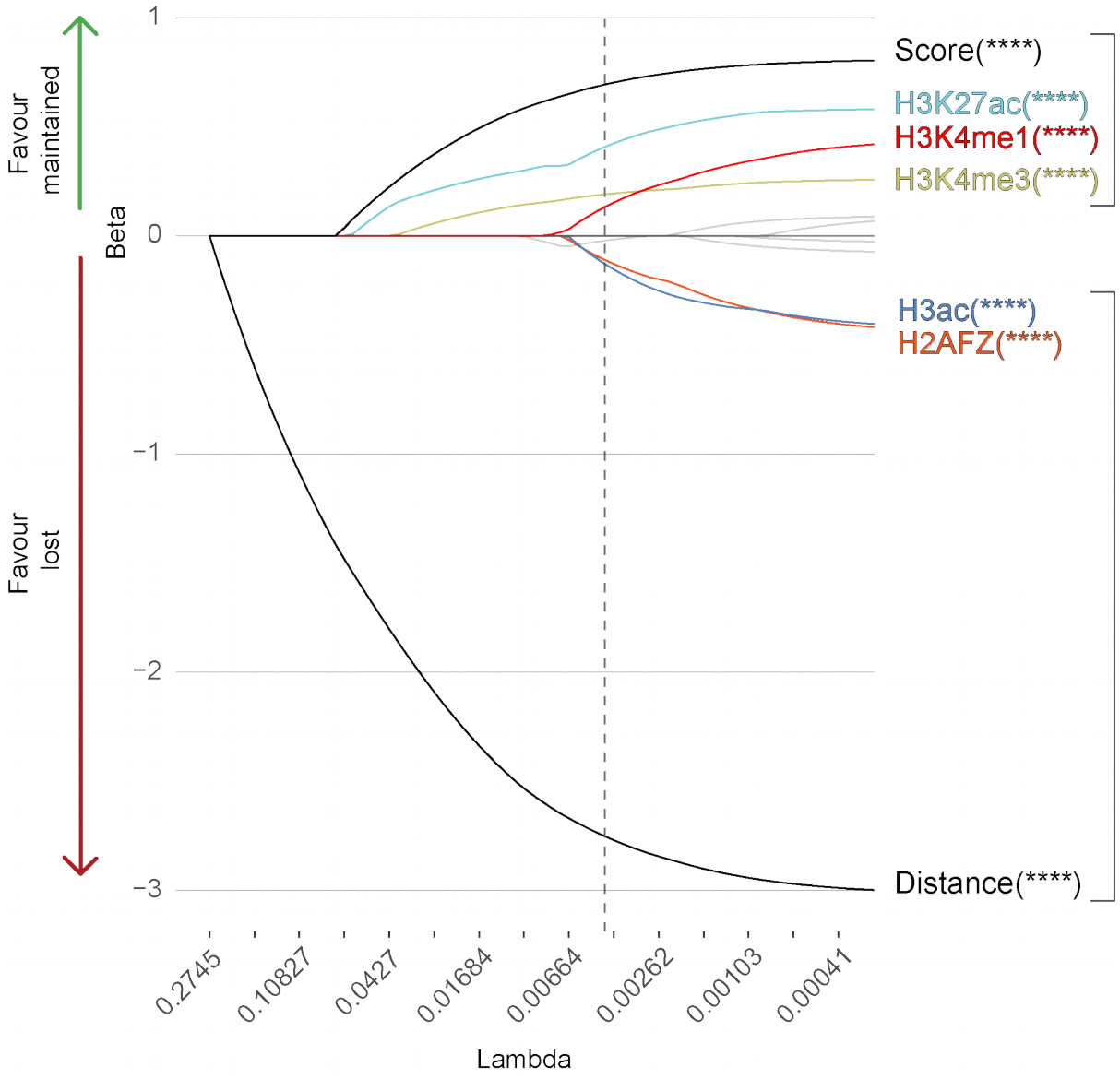

Figure 49. Lasso logistic regression regularization path of histones and histone modifications versus promoter interaction rewiring upon SCC1 depletion. The dependent binary variable represents maintained or gained promoter interactions. The independent variables comprise the ChIP-Seq scores for 11 histones and modifications thereof, in addition to the CHiCAGO score and the linear genomic

promoter interaction distance (log_{10}(bp)). The y-axis shows the regression beta which is positive for maintained promoter interactions and negative for lost promoter interactions. Note that gained promoter interactions are not taken into account in this analysis. The x-axis shows the lasso penalty coefficient, ranging from stringent on the left to permissive on the right. The vertical dotted line shows the lambda value at which the independent variables are tested for significance. H3K27ac, H3K4me1 and H3K4me3 are found to relate to promoter interaction maintenance. H3ac and H2AFZ are found to relate to promoter interaction loss. Asterisks indicate significance: **** $p < 0.0001$, *** $0.0001 < p \leq 0.001$, ** $0.001 < p \leq 0.01$, * $0.01 < p \leq 0.05$.

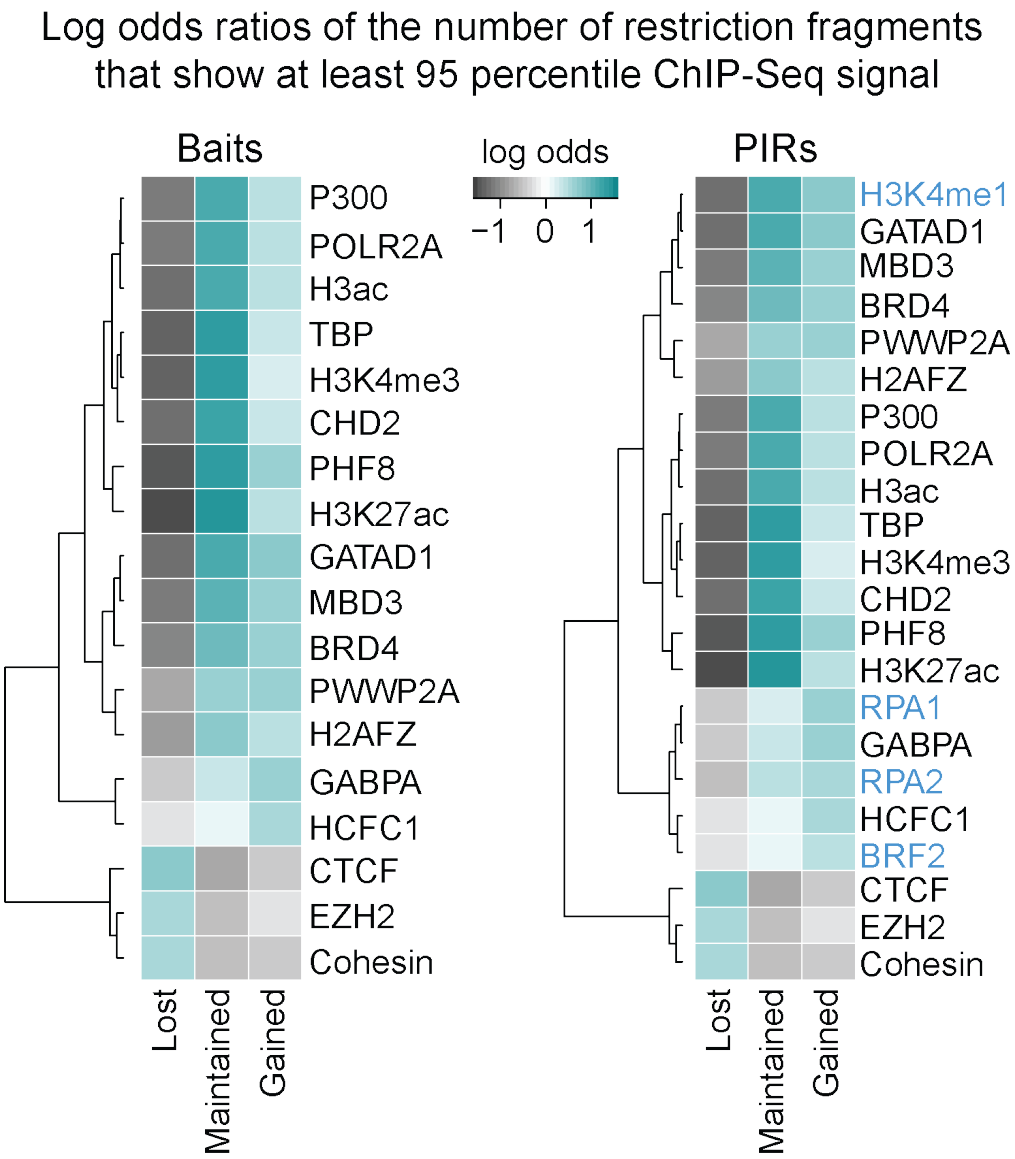

Figure 50. Strong ChIP-Seq signal overrepresentation at baits (left) and PIRs (right) involved in promoter interaction rewiring upon SCC1 depletion. The heat map shows natural log odds ratios of the number of RFs that display a ChIP-Seq signal strength of at least 95 percentile. Targets are only included in the heat map when the Fisher exact test is less than 0.05 (Benjamini Hochberg corrected). Hierarchical clustering was used to order the targets. Cohesin, CTCF and EZH2 significantly associate with RFs involved in lost promoter interactions. Four targets uniquely show significant differences at PIRs: H3K4me1, RPA1, RPA2 and BRF2 (shown in blue). The latter three cluster together with GABPA and HCFC1 (at baits and PIRs) and appear more strongly

overrepresented at RFs involved in gained promoter interactions than maintained.

4.4 Nascent transcription and chromatin accessibility in the context of promoter interaction rewiring upon SCC1 depletion

In the previous sections I show that promoter interactions can be categorized based on their response to SCC1 depletion. I then proceed to show how these categorized promoter interactions relate to interphase TAD organisation and which ChIP-Seq targets can be identified at their promoters and PIRs. In this section I aimed to relate promoter interaction rewiring to transcriptional response and to changes in chromatin accessibility. To detect transcriptional change upon SCC1 depletion, my collaborators performed SLAM-seq (Herzog *et al.*, 2017) experiments and data analysis (https://t-neumann.github.io/slamdunk/). SLAM-seq identifies nascent transcripts by means of detecting 4-thiouridine incorporation in newly synthesised RNA (Herzog *et al.*, 2017). This avoids detection of steady state RNA and provides data on newly differentially transcribed genes upon rapid SCC1 depletion.

The SLAM-seq experiment followed by differential expression analysis with DESeq2 revealed a total of 692 significantly differentially transcribed genes upon SCC1 depletion (DESeq2 FDR ≤ 0.05) (Figure 51). 421 genes were significantly up-regulated, 266 were significantly down-regulated, and 1197 genes showed little change in transcription (log2 fold change > -0.1 & log2 fold change < 0.1).

Selecting for strong misregulation upon cohesin depletion (log fold change ≥ 1.5 & FDR ≤ 0.05) yields 32 up- and 40 down-regulated genes. To ascertain that non-regulated genes were transcriptionally active, I obtained steady state RNA-Seq data from my collaborators, which I used to select transcribed non-regulated genes, leaving a total of 276 baits with active non-regulated genes (RPM ≥ 75th percentile).

Finally, note that assigning these data to baits inflates their numbers due to the fact that genes have multiple TSSs. This resulted in 649 baits of up-regulated genes, 337 baits of down-regulated genes, and 1536 baits of genes that show little transcriptional change.

To investigate whether the measured transcriptional changes upon cohesin depletion are related to promoter interaction rewiring of the involved genes, I first focused on strong misregulation (|log2 fold change| ≥ 1.5 & FDR ≤ 0.05). First, I asked whether the down-regulated genes were overrepresented among those genes that lose promoter interactions upon cohesin depletion. To test this, I express promoter

interaction rewiring and transcriptional response as a log odds ratio (see Methods 2.15). This ratio is > 0 if genes with lost promoter interactions tend to be more strongly down-regulated than genes with maintained or gained promoter interactions. I then performed this calculation on randomly sampled baits to construct an expected distribution of log odds ratios that shows what can be expected if there were no relation between promoter interaction rewiring and the transcriptional response. Figure 52 shows this expected distribution alongside the observed log odds ratio for lost promoter interactions. This reveals that strongly down-regulated genes upon cohesin depletion tend to lose promoter interactions (empirical p-value: ~0.029). I then performed the same analysis for the remaining combinations of regulation and rewiring: maintained and down-regulated, gained and down-regulated, lost and non-regulated, etc. Note that I used the same log2 fold change cut off of 1.50. I visualised the resulting z-scores as a heat map (Figure 53). Note that the bottom left heat map tile shows the same z-score as reported in Figure 52 (z = 1.88). These results reveal that upon cohesin depletion, active non-regulated genes show overrepresentation of maintained promoter interactions and that strongly up-regulated genes show overrepresentation of maintained and gained promoter interactions. In contrast, strongly down-regulated genes show overrepresentation of lost promoter interactions. Taken together, this analysis shows a systematic relationship between cohesin-dependent promoter interaction rewiring and transcriptional misregulation upon rapid cohesin depletion. Importantly, the fact that non-regulated genes show overrepresentation of maintained promoter interactions suggests that cohesin-independent promoter interactions may facilitate transcriptional robustness to cohesin depletion. Lastly, it should be noted that the observed relationship is undetectable when including weaker differentially transcribed genes in the analysis, presumably as a result of a lower signal to noise ratio.

I next sought to further investigate the relationship between promoter interaction rewiring upon cohesin depletion and transcriptional response. Notably, I aimed to integrate the SLAM-Seq data without imposing a stringent cut off on the log fold change. Furthermore, I sought to include epigenetic signatures in the analysis. To this end I asked whether promoter interaction rewiring can be used as an independent variable in a regression analysis to predict transcriptional change in the context of cohesin depletion, and if so, if the performance improves when selecting

only interactions between promoters and putative enhancers. To this end, I performed an ordinal logistic regression analysis, since this provides a formal statistical method for analysis of the ordinally stratified dependent variable: up-, non- and downregulated (1, 0, -1 respectively). Here, I used a less stringent log2 fold change cut off of 0.1, which results in 589 baits of up-regulated genes, 304 baits of down-regulated genes and 276 baits of active non-regulated genes. As the independent variable, I used the relative change in promoter wiring for each gene (see Methods 2.13.4). A large value of the independent variable corresponds to a bait with more gained than lost and maintained promoter interactions. Conversely, a small value of X corresponds to a bait with more lost than gained and maintained promoter interactions. I constructed this independent variable twice; once using PIRs bearing epigenetic marks of activity (H3K4me1, H3K4me3 and H3K27ac) and once using PIRs that lack all three of those marks. Since many PIRs do not meet these criteria, and because many genes that show significant transcriptional response upon cohesin depletion do not show detectable promoter interaction rewiring, I was left with 426 baits after integrating the dependent and independent variables (93 up-regulated, 189 non-regulated, and 144 down-regulated). I then performed ordinal logistic regression separately on the active and the non-active sets, the results of which are shown in Figure 54. This shows that upon cohesin depletion, genes with gained promoter interactions tend to be up-regulated while genes with lost promoter interactions tend to be down-regulated. However, this relation is only significant when the rewiring promoter interactions are with PIRs that bear active epigenetic marks ($p \approx 8.63*10\text{-}3$). Furthermore, promoter interaction gain or loss does not appear to correlate with genes that show no transcriptional response upon cohesin depletion, independently of the presence of active epigenetic marks at the PIRs. These results are in agreement with the previous analysis (Figure 53): upon cohesin depletion, transcriptional up-regulation is related to gained promoter interactions while transcriptional down-regulation is related to lost promoter interactions. Furthermore, concordant with the findings of (Javierre *et al.*, 2016), active epigenetic marks appear to be a relevant property for predicting transcriptional response. This suggests that promoter enhancer interactions are a key mechanism that drives transcriptional change upon cohesin depletion.

Since two components of the cohesin complex (RAD21 and STAG1) are likely involved in RNApolII specific transcriptional regulation (Liu *et al.*, 2009), it is possible that the observed transcriptional changes are driven by loss of cohesin in cis. To investigate whether promoter interaction rewiring upon cohesin depletion imbues greater influence on transcriptional activity than loss of cohesin at the promoter, I focused on active enhancers that contact multiple promoters, and whose promoter interactions respond differently to cohesin depletion. E.g. in the presence of cohesin, one PIR interacts with the promoters of two genes while upon cohesin depletion, one gene maintains its promoter interaction with the PIR, whereas the other loses its promoter interaction. I reasoned that if the transcriptional activity of both genes is reduced despite only one of them losing the interaction, this would be consistent with effects of cohesin depletion, unrelated to promoter interaction rewiring. Conversely, if only the gene that loses its promoter interaction with the enhancer is downregulated, while the gene that retains it is unaffected, this suggests that the promoter interaction with this enhancer may drive gene transcription. I used the compendium of ChIP-Seq data compiled as described in section 2.10 to select PIRs with features of enhancers (H3K4me1, K3K4me3 & H3K27ac), and with the rewiring properties specified above. I found a total of 15 enhancer-containing PIRs with the required rewiring properties. In the presence of cohesin these enhancers contact the promoters of 34 genes (with 4/15 PIRs contacting 3 promoters). Upon cohesin depletion, 15 promoter interactions were lost, while 19 were maintained. Figure 55 shows the transcriptional response of the corresponding genes. This reveals that transcription is lower when the interaction with the shared enhancers is lost than when it is maintained ($p = 0.048$, see Methods 2.16). Taken together, this analysis is consistent with the model that gene expression changes upon cohesin depletion are triggered by the loss of cohesin-specific connections between enhancers and promoters.

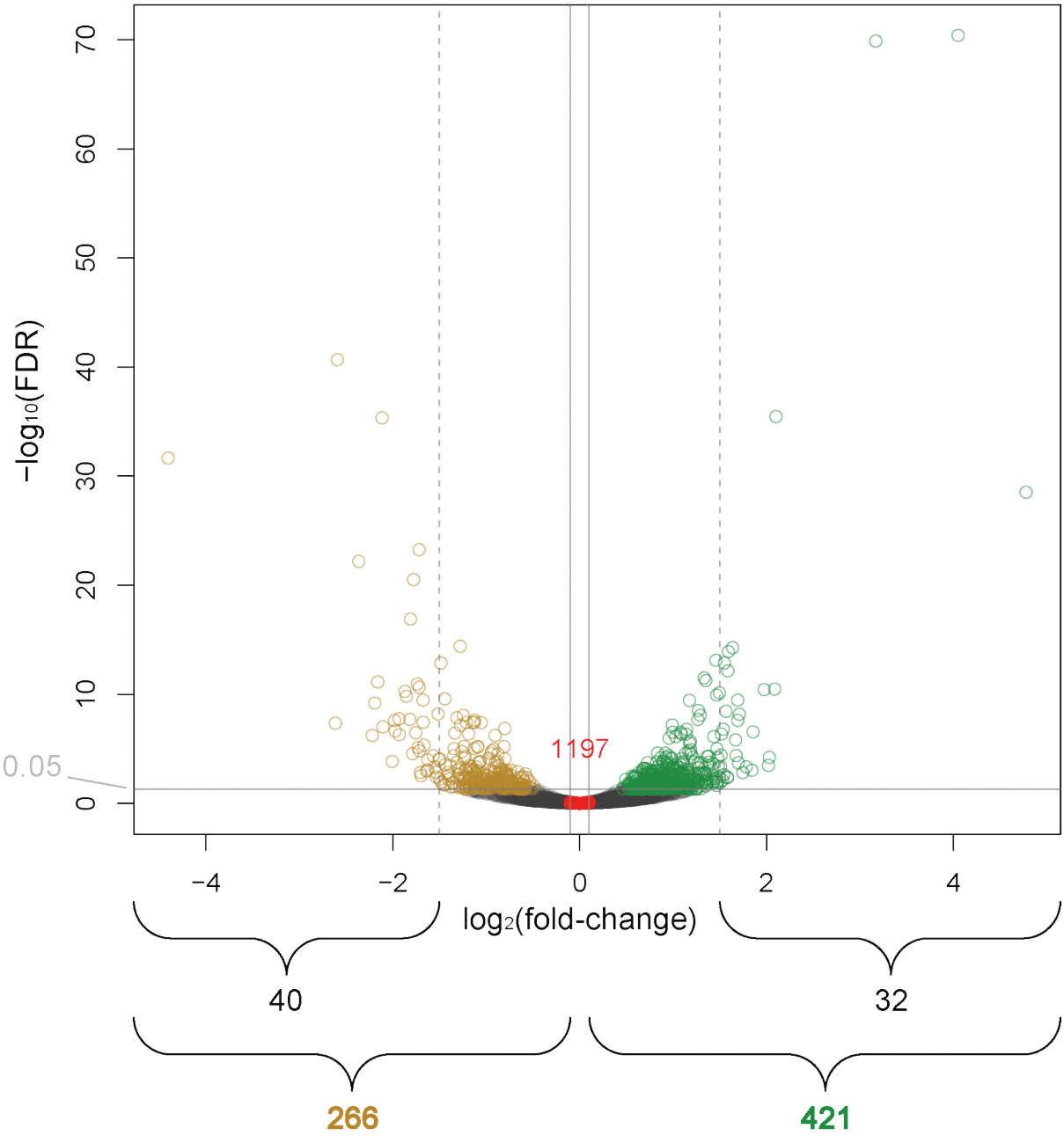

Figure 51. SLAM-Seq nascent RNA levels upon SCC1 depletion. In total, transcripts were detected for 7424 genes. I categorised genes as: down-regulated, non-regulated and up-regulated based on a FDR cutoff of 0.05 and the $\log_2$ fold-change. I used a stringent (1.5) and a more permissive (0.1) fold-change cutoff. With a fold-change cutoff of 1.5, I find 40 down-regulated genes and 32 up-regulated genes. With a fold-change cutoff of 0.1, I find 266 down-regulated genes and 421 up-regulated genes. Note that the number of non-regulated genes (1197) is not

affected by the stringency. However, using RNA-Seq data, I have made a sub-selection of *active* non-regulated genes, resulting in 276 genes.

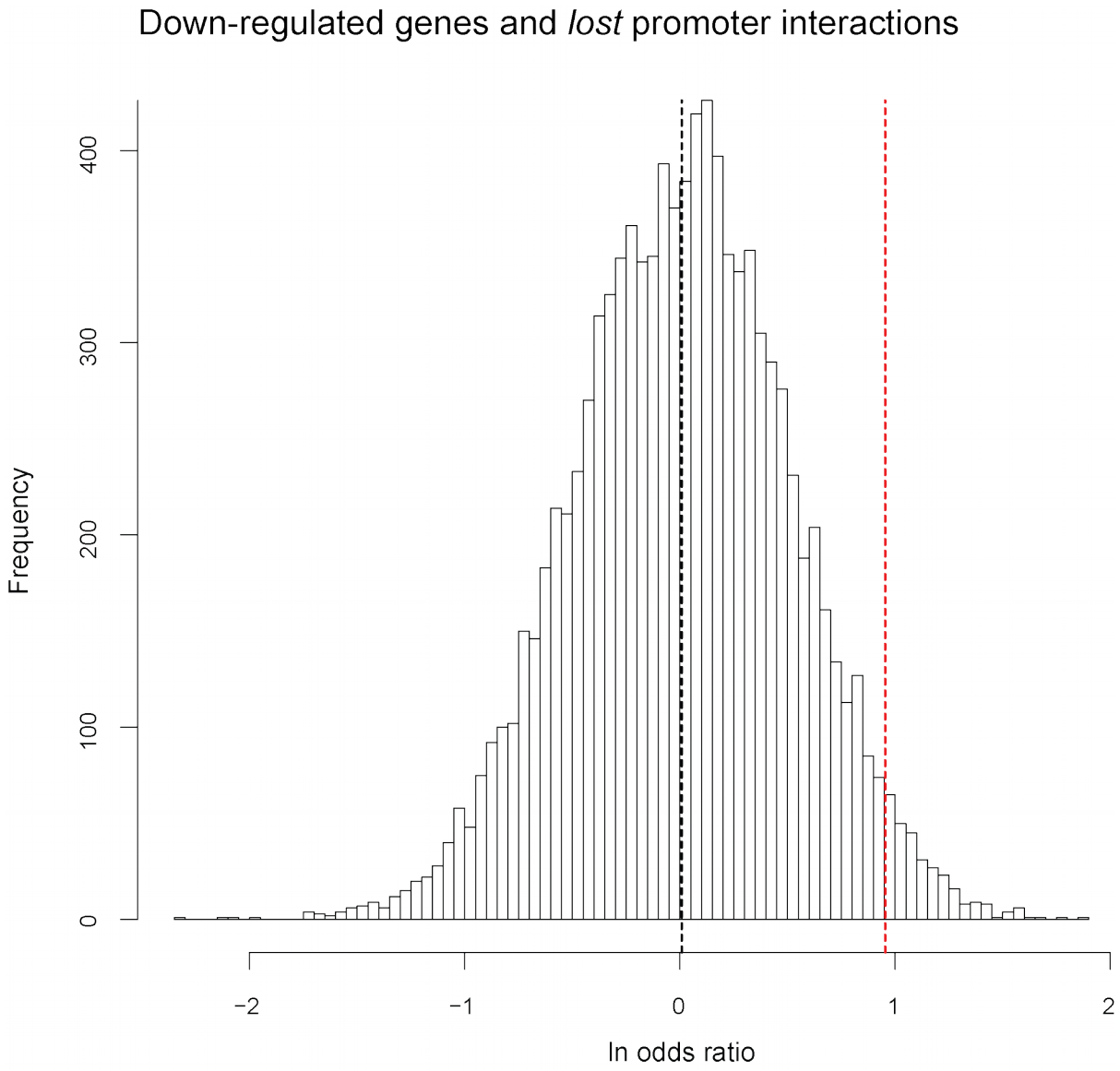

Figure 52. Strong transcriptional down-regulation upon SCC1 depletion coincides with promoter interaction loss. The red dotted line shows the observed log odds ratio of the number of promoter interactions of strongly down-regulated genes versus not down (i.e. up-regulated or non-regulated) versus their rewiring: *lost* versus not *lost* (i.e. *maintained* or *gained*). The expected distribution is the result of performing this analysis on randomized baits, 10,000 times. The observed log odds ratio is greater than the majority of expected values, indicating that down-regulated genes tend to lose promoter interactions (empirical p-value = 0.029, z-sore = 1.88).

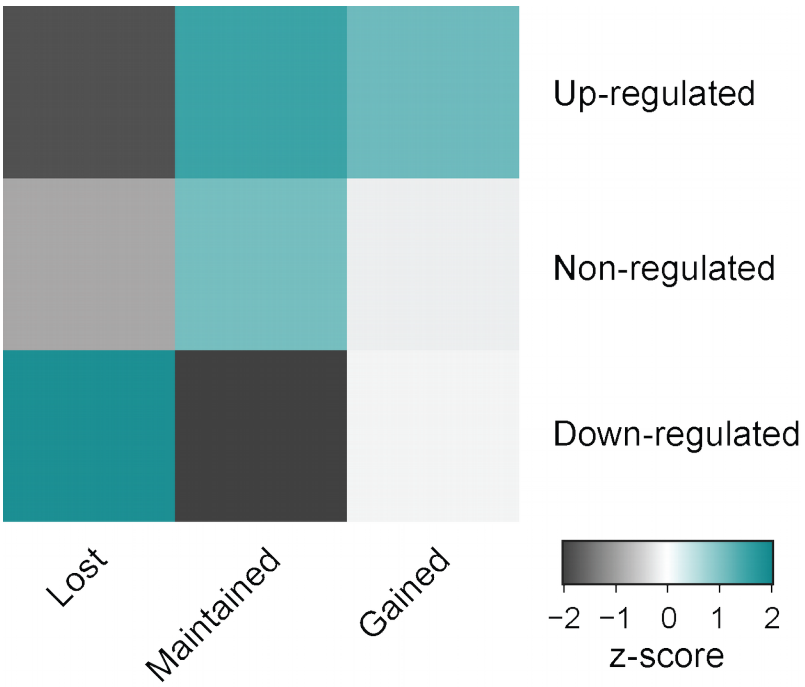

Figure 53. Transcriptional response upon SCC1 depletion coincides with lost, maintained, and gained promoter interactions. This heat map shows the log odds ratios from nonparametric bootstrap analyses expressed as a z-score (as in Figure 52). This shows that transcriptionally strongly up-regulated genes tend to have promoter interactions that are maintained and gained upon SCC1 depletion. Furthermore, genes that show little transcriptional change tend to have promoter interactions that are maintained. Lastly, strongly down-regulated genes show a tendency to have promoter interactions that are lost and conversely, tend not to have maintained promoter interactions.

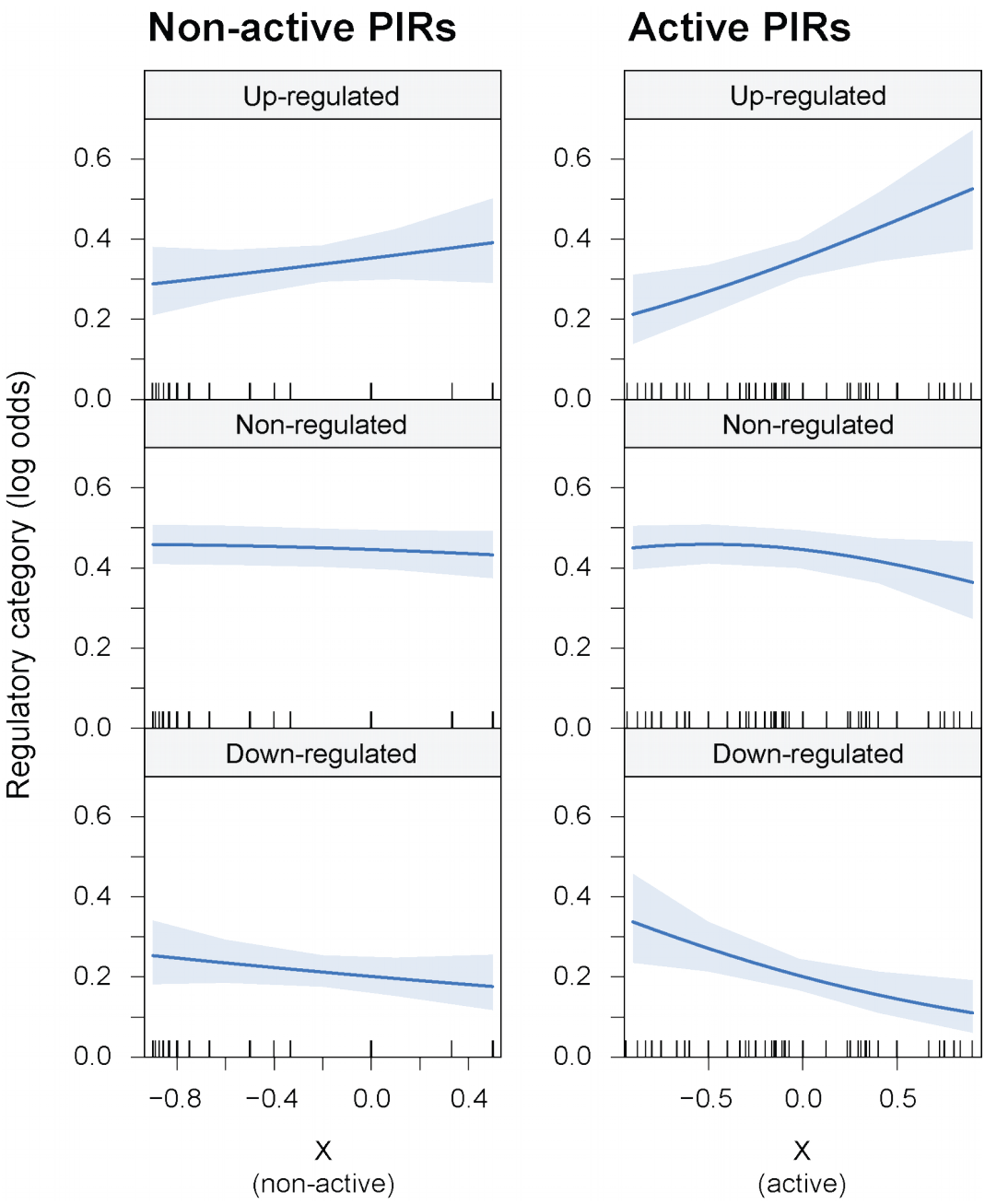

Figure 54. Transcriptional response to cohesin depletion is related to rewiring of promoter interactions with active PIRs. Effects plot of ordinal logistic regression on cohesin-depletion-dependent nascent transcriptional activity and promoter interaction rewiring. The ordinal dependent variable represents transcriptional response upon cohesin depletion: up-regulation, non-regulation and down-regulation. The independent variable (X) represents promoter interaction rewiring with PIRs bearing epigenetically active marks H3K4me1, H3K4me3, H3K27ac (right) and those that lack these marks (left). The independent variable

describes the difference between the number of gained and lost promoter interactions per bait as a proportion of all interactions per bait (see Methods 2.13.4). The regression shows that genes that gain active promoter interactions tend to be up-regulated while genes that lose active promoter interactions tend to be down-regulated (p = 8.63*10-3). This relation is not significant for promoter interaction rewiring with PIRs that do not bear active marks.

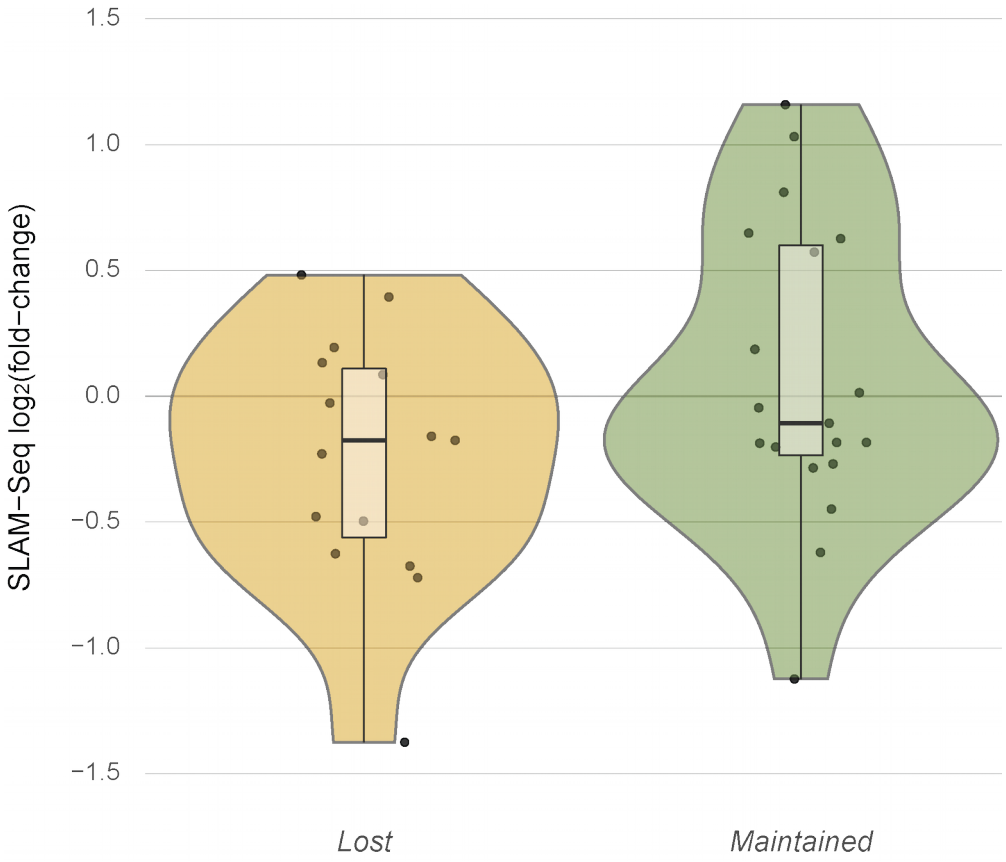

Figure 55. Genes that share differentially rewiring PIRs upon cohesin depletion show different transcriptional responses. SLAM-seq log2 fold change upon cohesin depletion of genes with lost promoter interactions (left) and maintained promoter interactions (right). Only PIRs bearing active epigenetic enhancer marks were taken into account. Additionally, these PIRs contact multiple genes. Upon cohesin depletion the promoter interaction with at least one gene is lost while the promoter interaction with at least one other gene is maintained. Using these criteria, 15 enhancers were found which interacted with promoters of 34 genes (4 enhancers contacted 3 promoters). 15 promoter interactions

were lost while 19 were maintained. Genes that lose promoter interactions with such PIRs are significantly more down-regulated upon cohesin depletion than those that maintain the interactions ($p = 0.048$).

I next asked whether cohesin depletion results in differences in the chromatin accessibility of enhancers *in cis*. Since chromatin accessibility driven by compaction/ decompaction is related to transcriptional regulation (Gribnau *et al.*, 2000), it is possible that this mechanism operates in conjunction with promoter interaction rewiring to produce the observed transcriptional response upon cohesin depletion. To investigate this, together with Dr Valeriya Malysheva, I performed ATAC-seq analyses on SCC1-AID -Auxin, SCC1-AID +Auxin, CTCF-AID -Auxin, and CTCF-AID +Auxin HeLa cells. ATAC-seq uses the transposase TN5 to fragment the genome while integrating sequencing adapters. This integration has a higher propensity for open chromatin, therefore a higher ATAC-seq signal correlates with less occupied DNA. After aligning and quality control, during exploratory analyses, I found that principal components 1 and 4 (24% and 4% of variance) separated the samples according to the conditions (Figure 56). The SCC1 samples showed the strongest separation along the first principal component. This suggests a possible relationship between cohesin depletion and differential chromatin accessibility. To further investigate this, I devised a strategy to combine TN5 transposase integration site counts directly with RFs (see Methods 2.13.5). This approach exploits the base-pair resolution that ATAC-seq provides by moving away from the read-level to the level of exact integration sites. I then performed DESeq2 analysis on the TN5 integration counts per RF upon SCC1 depletion. This resulted in 3593 significantly differentially compacted RFs (FDR ≤ 0.05), 941 of which showed increased TN5 integration and 2652 showed decreased TN5 integration upon SCC1 depletion (see Figure 57).

I first asked whether the observed changes in TN5 integration occur at the promoters of misregulated genes. To this end, I compared the SLAM-seq results per bait to the ATAC-seq results. I visualised the nascent transcriptional response (SLAM-seq) upon cohesin depletion of genes with baits that show a significant change in TN5 integration (FDR ≤ 0.05) upon cohesin depletion (616 “compacted”, 61 “de-compacted”) (Figure 58). This clearly shows that “compacted” baits correspond to genes that are transcriptionally down-regulated upon cohesin depletion, while “de-compacted” baits correspond to genes that are transcriptionally up-regulated upon

cohesin depletion. On the whole, this result indicates a role for differential accessibility at promoters in the context of transcriptional misregulation upon cohesin depletion.

I next investigated whether differential TN5 integration at PIRs might explain, at least in part, the transcriptional response upon cohesin depletion. To this end, I performed an ordinal logistic regression analysis, using TN5 integration scores per RF. I processed these similarly to the previously shown ChIP-Seq scores per RF (see Methods 2.11.2). The dependent variable was the SLAM-Seq regulatory category: up-, non-, or down-regulated (see Figure 57). I constructed an independent variable (dcv) that describes differential “compaction” in the context of promoter interaction rewiring. I.e. if a promoter interaction with an “open” PIR is lost, this negatively contributes to the independent variable score. Conversely, if a promoter interaction is maintained and the PIR is “decompacted” upon cohesin depletion, this positively contributes to the independent variable score. However, if a promoter interaction is maintained but the PIR is “compacted” upon cohesin depletion, this negatively contributes to the independent variable score. Lastly, if a promoter interaction is gained this contributes positively if the PIR is “open” but negatively if the PIR is “compacted” (see Methods 2.13.5 for a formal description of the independent variable). In short, the independent variable describes the joint effect of promoter interaction rewiring and differential “compaction” at the PIRs. Since I detect few genes that simultaneously show significant transcriptional change as well as promoter interaction rewiring, and since not all PIRs can be reliably categorised into “compacted”, “no change” and “de-compacted” (Figure 57), combining these data results in a reduction of the total number of baits (607 baits remained). I performed ordinal logistic regression using these data, the results of which are shown in Figure 59. This shows that there is a positive relationship between transcriptional regulation upon cohesin depletion and the joint rewiring and differential compaction variable ($p = 0.016$ for a negative value of the independent variable X, and $p = 0.005$ for a positive value of X). This shows that differential chromatin compaction at PIRs can explain the transcriptional response upon cohesin depletion.

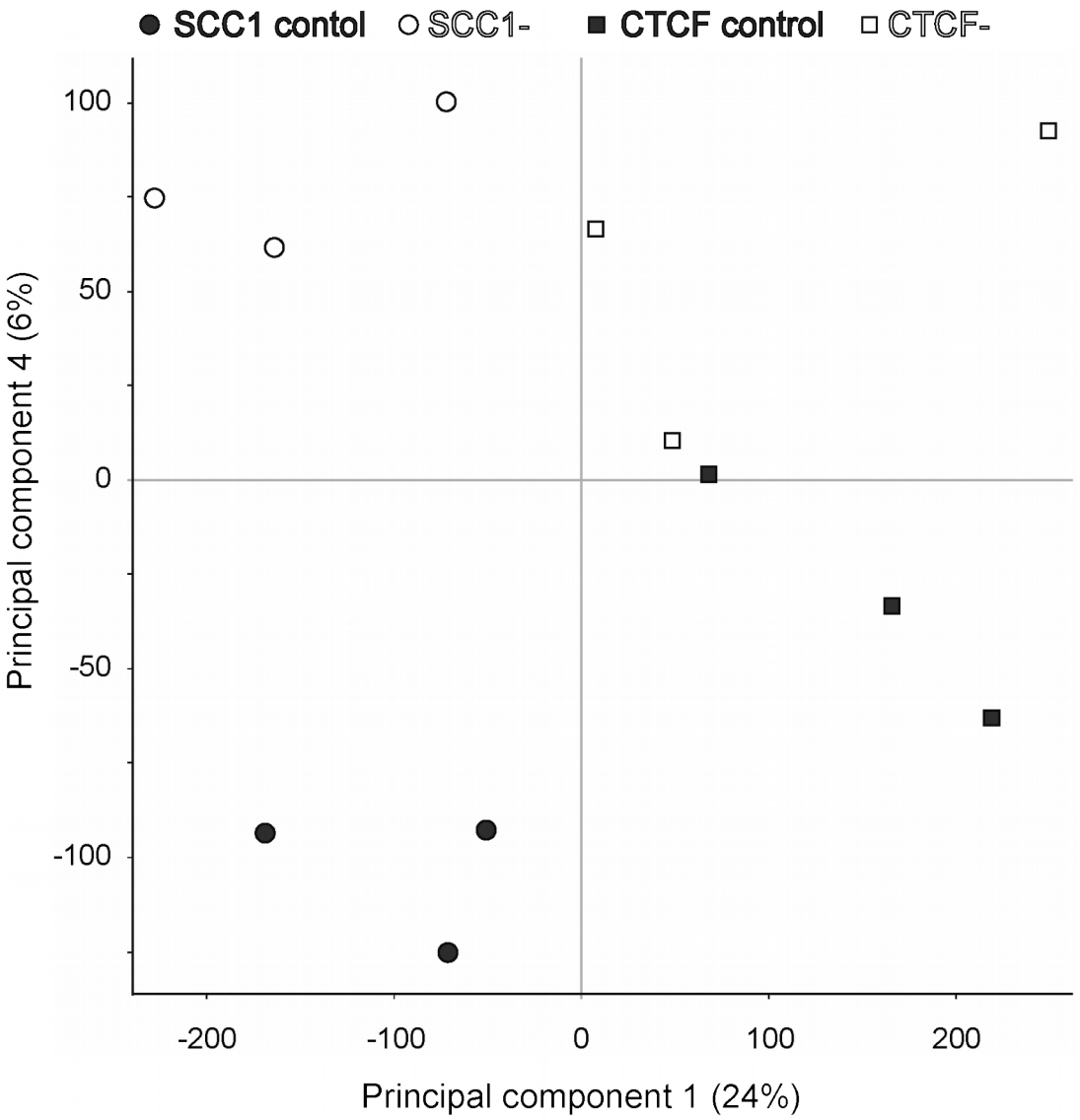

Figure 56. Principal component analysis of ATAC-Seq data suggests cohesin-depletion-dependent differential chromatin compaction. PCA of binned ATAC-Seq reads of SCC1-AID +/-Auxin and CTCF-AID +/-Auxin (10kb bins, 5kb step). Samples segregate into groups according to conditions when observing components 1 and 4. These components encompass 24% and 6% of the variance respectively. The SCC1 +Auxin and -Auxin samples separate clearly along the fourth principal component (y-axis), whereas the CTCF samples separate to a lesser extent.

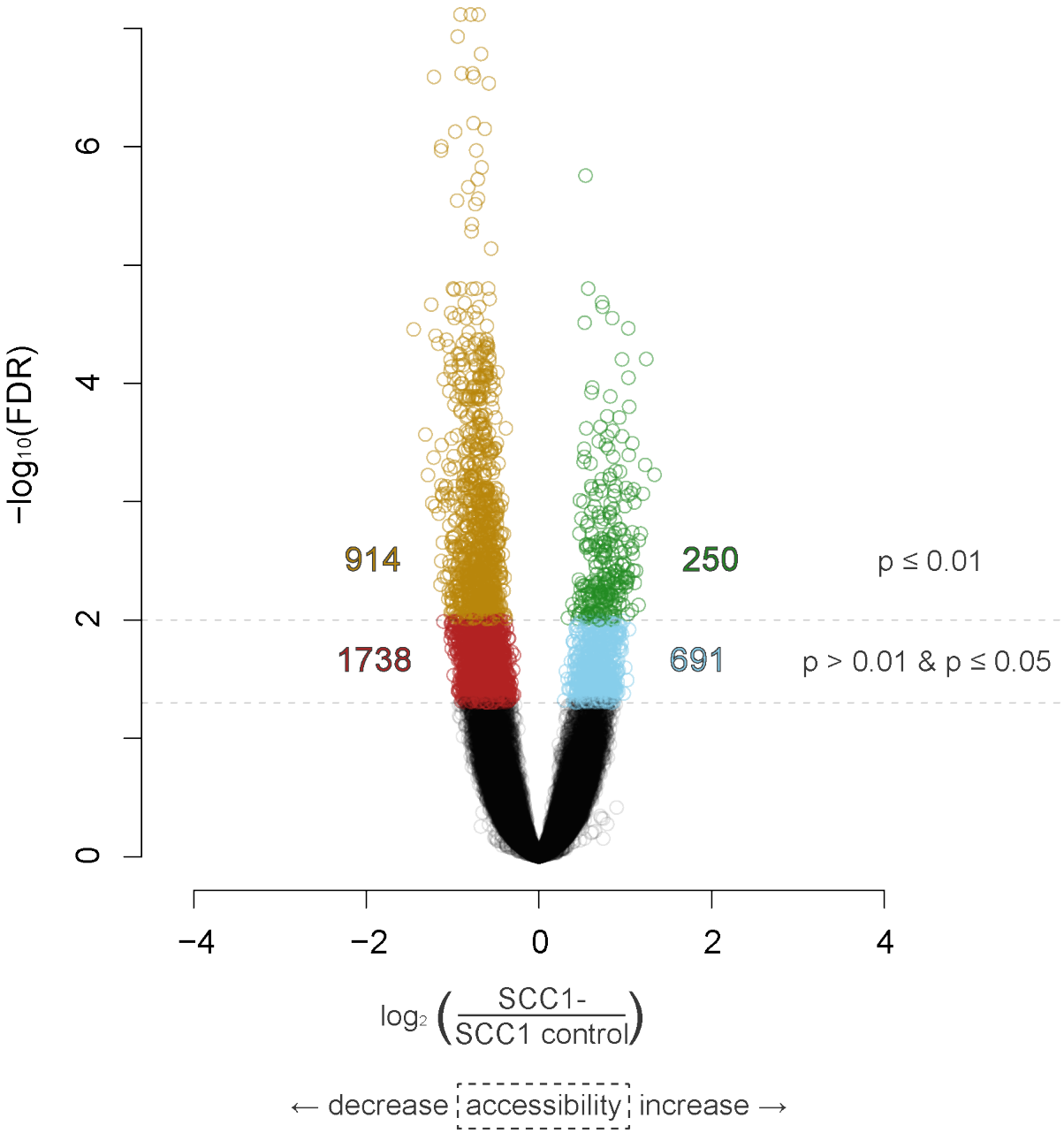

Figure 57. Changes in accessibility detected upon cohesin depletion. Volcano plot of differential ATAC-Seq TN5 transposase integration events per RF detected with DESeq2. The x-axis shows the ATAC-seq log2 fold-change score between the depleted and the control conditions. A higher log2 fold-change corresponds to an increase in accessibility upon SCC1 depletion. The y-axis shows the -log10 FDR, with cut-off values indicated as dashed lines at FDR = 0.05 and FDR = 0.01. 3593 significantly differentially compacted RFs are detected (FDR ≤ 0.05), of which 2652 show decreased accessibility (yellow and red) and 941 show increased accessibility (green and blue) upon SCC1 depletion.

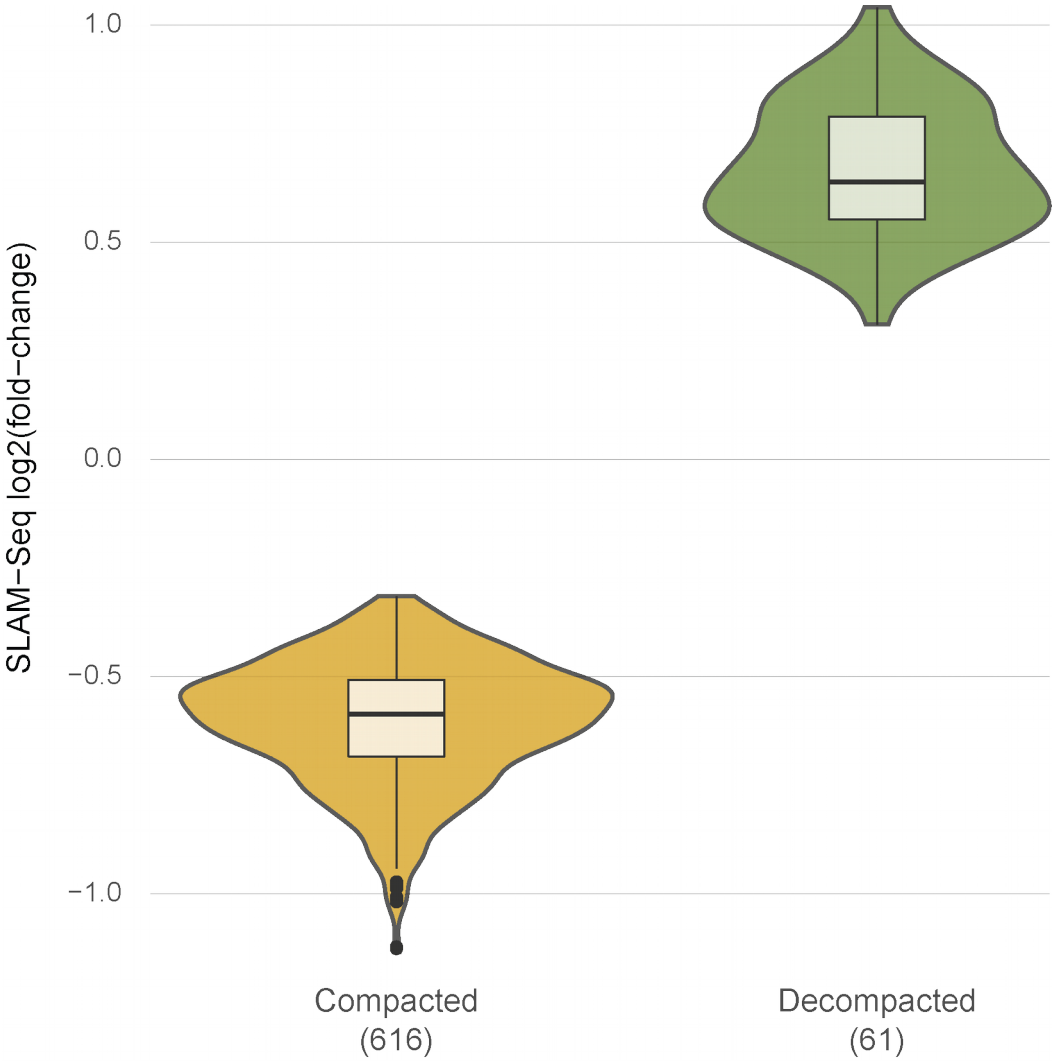

Figure 58. Differential compaction at promoters and transcriptional change upon cohesin depletion are related. Significantly differentially compacted baits (FDR <= 0.05) show clear differences in nascent transcriptional activity: compacted baits show transcriptional down-regulation, while de-compacted baits show transcriptional up-regulation. 649 baits show compaction upon cohesin depletion whereas only 63 baits show decompaction upon cohesin depletion.

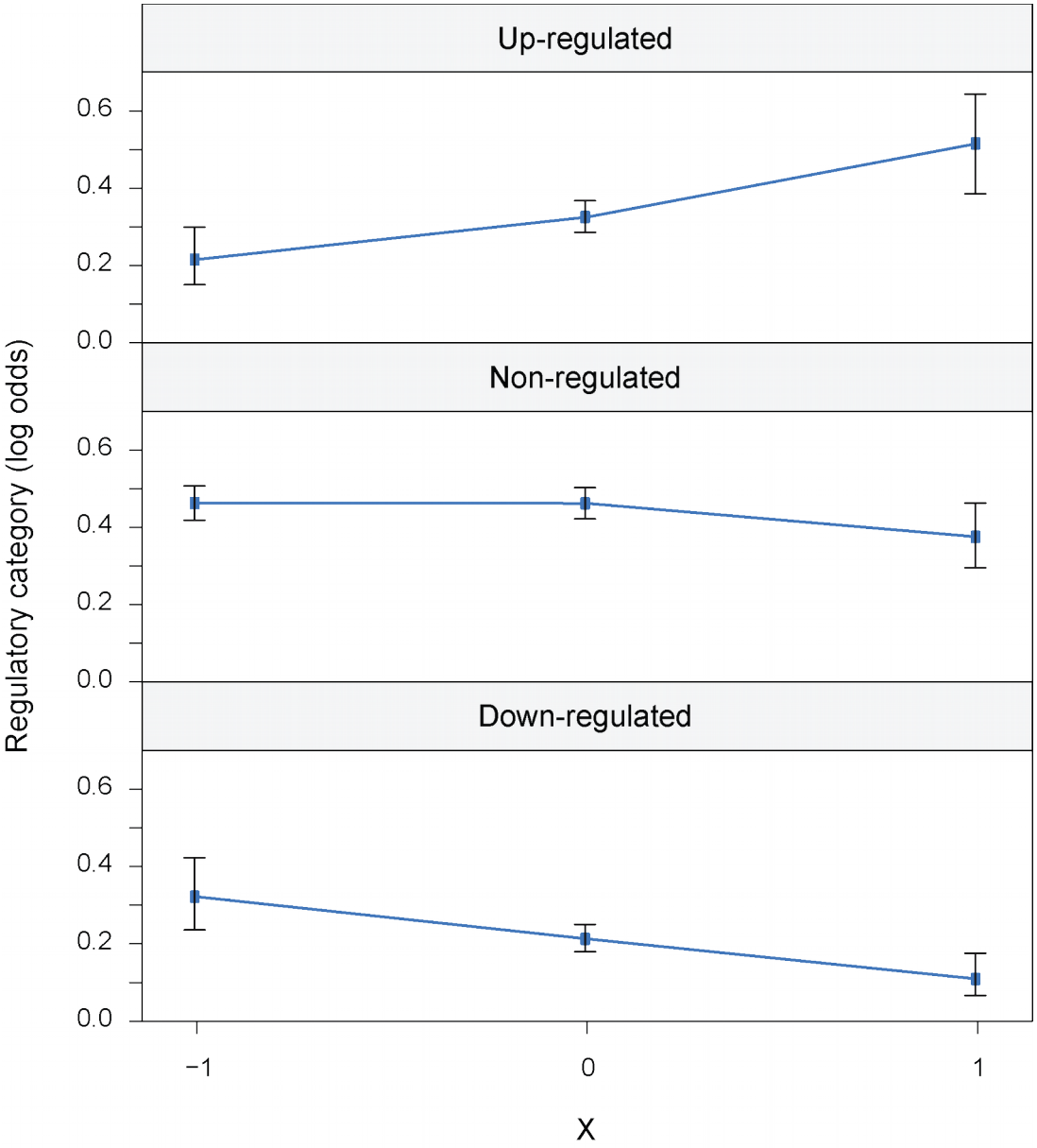

Figure 59. Transcriptional response to cohesin depletion is related to promoter interaction rewiring with differentially compacted PIRs. Ordinal logistic regression of transcriptional response upon SCC1 depletion (SLAM-Seq) versus chromatin compaction response at PIRs upon SCC1 depletion. The ordinal dependent variable represents transcriptional response upon cohesin depletion: up-regulation, non-regulation and down-regulation. The independent variable (X) represents promoter interaction rewiring with PIRs that show differential compaction upon cohesin depletion (see Methods 2.13.5). The independent variable shows a positive beta at 1 (0.79) and a negative beta at -1 (-0.56). The respective p-values are ~0.005 and ~0.016. This result indicates that

chromatin accessibility at PIRs changes concurrently with the transcriptional response upon cohesin perturbation.

4.5 Validation of the functional effects of cohesin-dependent and independent promoter interactions

To confirm whether cohesin dependent promoter interaction rewiring imbues transcriptional change, together with Drs Wutz, Tang (JM Peters lab) and Spivakov, I devised two validation experiments: perturbation of transcriptional activity with subsequent PCHi-C analysis, and perturbation of putative enhancers in the context of cohesin dependent promoter interaction rewiring.

The first validation experiment aims to verify that promoter interactions are supported by transcriptional machinery. I expect that promoter interactions that are lost upon transcriptional perturbation to frequently be the same as those that I categorized as maintained upon SCC1 depletion. The RNApolII inhibitor Triptolide was selected to globally perturb transcriptional activity. Triptolide inhibits transcriptional activity by inducing degradation of Pol II subunit Rbp1 (Wang *et al.*, 2011). Dr Wen Tang has performed the Triptolide experiments. Subsequent PCHi-C experiments are currently being performed by Dr Steven Bevan (Stefan Schoenfedler's lab).

The second validation experiment aims to validate the hypothesis that enhancers engaged in either cohesin-dependent or cohesin-independent interactions with their target promoters are relevant for transcriptional control. To this end, we use CRISPRi dCAS9-KRAB, targeted at putative enhancer regions that I identified using PCHi-C, ChIP-Seq and ATAC-seq. KRAB (Kruppel-associated box) functions as a repressor by inducing chromatin compaction (Gilbert *et al.*, 2013). The aim was to use this property of chromatin compaction to silence enhancers in a precise and targeted way. I selected five genes based on transcriptional response upon cohesin depletion: two strongly down-regulated (NUAK1 and SLC16A6) and three non-regulated (BAHCC1, TRIM66 and PDE3A). I used RNA-Seq to determine steady state transcriptional activity of these genes. I selected the down-regulated genes on presence of lost promoter interactions. Conversely, I selected the non-regulated genes on presence of maintained promoter interactions. I narrowed down the locations of the enhancers within the PIRs by using ChIP-Seq data on cohesin, H3K27ac, H3K4me1, and H3K4me3, in addition to ATAC-seq data. Namely, I

prioritized regions that showed presence of enhancer marks H3K27ac and H3K4me1 alongside ATAC-Seq peaks, while showing absence of the canonical promoter mark H3K4me3. Furthermore, I prioritised regions that showed heightened cohesin presence. The resulting fine-mapped CRISPRi targets ranged from 115bp to 884bp, with a mean of ~403.6bp. As an illustrative example, Figure 60 shows the selected putative enhancer targets of the strongly down-regulated gene NUAK1 (vertical red bars). All but one of these selected putative enhancers show MACS2 peaks for both H3K27ac and H3K4me1, two show ATAC-Seq peaks and none show H3K4me3 peaks. Furthermore, three out of four interphase TAD samples show that the promoter of NUAK1 and its interactions are located within the same TAD. Finally, none of the selected putative enhancers are located at known promoter regions of other genes.

Upon targeting dCAS9-KRAB to the putative enhancers (in the presence of cohesin), RT-qPCR shows decreased transcript abundance for NUAK1 and SLC16A6 (Figure 61). This indicates that the lost promoter interactions of these genes were indeed with enhancers which support transcriptional activity of these genes. Targeting the core promoter regions of these genes leads to slightly more reduced transcript levels, suggesting that silencing the enhancers leads to considerable transcriptional silencing. Upon targeting the putative enhancers of BAHCC1 and TRIM66 we also detected decreased transcript abundance, although the extent was limited and the variance was large. However, targeting the promoter regions of these genes showed a similarly weak reduction in transcript levels. This suggests that these genes may have been transcriptionally near-silent under control conditions. Alternatively, the transcripts of these genes may show low turnover, leading to retention and subsequent detection of steady state mRNA. BAHCC1 and TRIM66 showed no SLAM-seq signal which can indicate absence of transcriptional activity, although this is not conclusive due to the sparsity of the dataset. Upon inspection of the HeLa epigenome (Zerbino *et al.*, 2015), the promoter region of BAHCC1 appears to lack active marks, supporting the notion that this gene may show little transcriptional activity.

Taken together, these preliminary results indicate that transcriptional down-regulation upon cohesin depletion appear to be caused by loss of promoter-enhancer interactions. In contrast, the results on maintained promoter interactions

are currently inconclusive. I envision that the ongoing validation experiments will provide valuable insights into this matter.

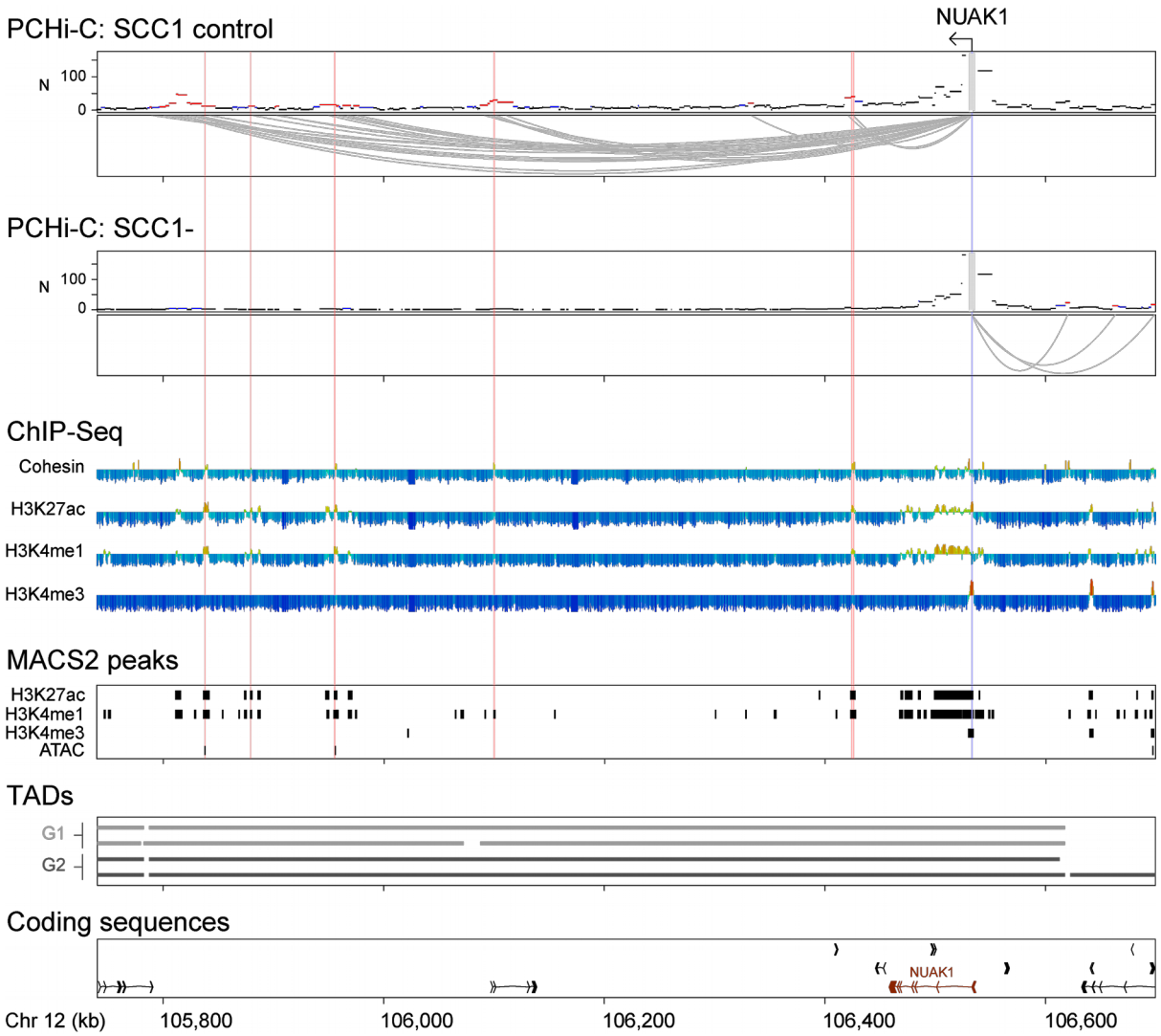

Figure 60. Selection of enhancer regions for CRISPRi dCAS9-KRAB perturbation. The enhancers interact with the promoter of NUAK1. From top to bottom: PCHi-C interactions under control SCC1 -Auxin conditions, PCHi-C interactions under depleted SCC1 +Auxin conditions, ChIP-Seq tracks (Cohesin, H3K27ac, H3K4me1, H3K4me3) MACS2 peaks (H3K27ac, H3K4me1, H3K4me3, ATAC-Seq), TAD intervals, and lastly coding sequences. Note that the cohesin ChIP-Seq data derive from anti-SCC1 analyses. PIRs are selected for validation with CRISPRi dCAS9-KRAB based on presence of enhancer marks H3K26ac and H3K4me1. Target promoter interactions are shown as vertical red bars. The vertical blue bar represents the targeted promoter regions.

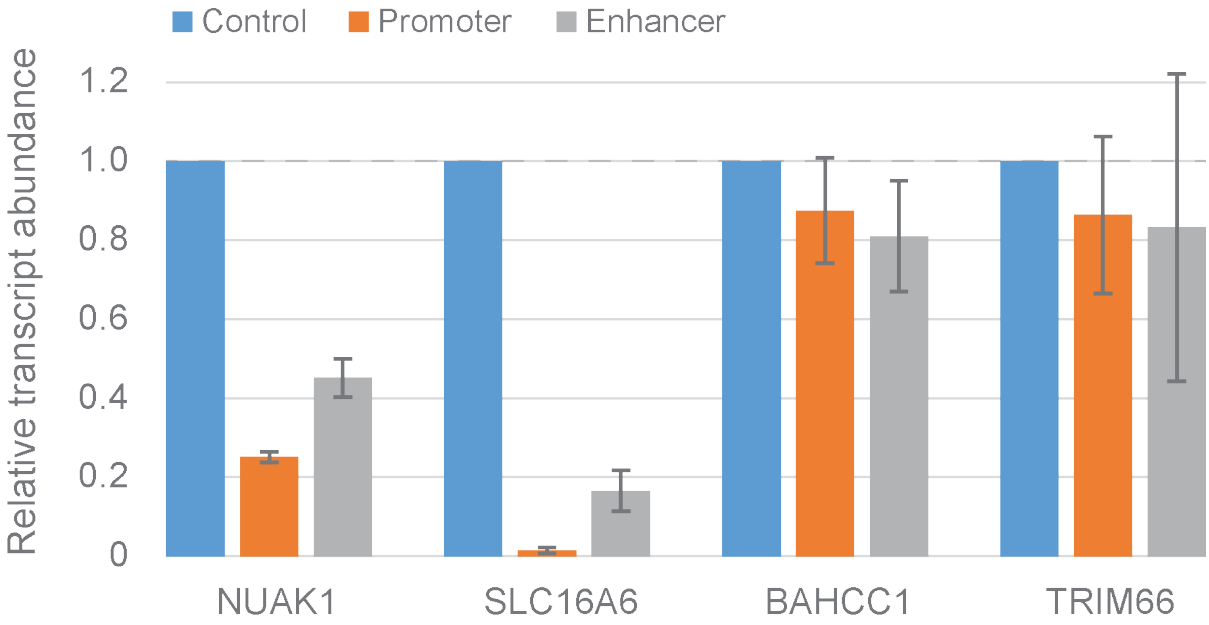

Figure 61. Inactivating enhancers recapitulates transcriptional loss upon promoter interaction loss. Preliminary RT-qPCR results for CRISPRi-dCAS9-KRAB targeted at putative enhancers of NUAK1, SLC16A6, BAHCC1 and TRIM66. Bar height shows mean transcript abundance (n = 2), error bars show mean +/-1SE. Transcript abundance has been normalised relative to the housekeeping gene GAPDH. The control consists of a CRISPRi experiment lacking a gRNA. NUAK1 and SLC16A6 show strong decreases in transcript abundance upon perturbation of putative enhancers as well as upon targeting the core promoter region. Although BAHCC1 and TRIM66 show mean transcript level decrease upon targeting of putative enhancers and core promoters, variance is high. NUAK1 and SLC16A6 are genes with cohesin dependent promoter interactions that lose transcriptional activity upon cohesin depletion. BAHCC1 and TRIM66 are genes with cohesin independent (maintained) promoter interactions that do not lose transcriptional activity upon cohesin depletion.

5 Discussion

5.1 TAD-boundary crossing

5.1.1 The capacity of TAD boundaries to constrain promoter interactions

In chapter 3, I show that TAD boundaries appear to insulate promoter interactions imperfectly, a feature that appears consistent between various human tissues. I integrated TAD partitionings detected in Hi-C data with “focused” promoter interactions detected in PCHiC data to evaluate the relationship between these two layers of chromosomal organisation. I show that ~30% of all promoter interactions cross TAD boundaries in eight human haematopoietic lineages. This result, now published as part of (Javierre *et al.*, 2016), is supported by more recent additional reports that observe inter-TAD DNA interactions (Cattoni *et al.*, 2017; Flyamer *et al.*, 2017; Bintu *et al.*, 2018). Additionally, I show that cell cycle synchronised HeLa cells exhibit similar proportions of TAD boundary crossing promoter interactions. Lastly, I identified promoters that interact solely across TAD boundaries and provide evidence that these promoter interactions may be involved in transcriptional regulation. Together, these findings suggest that TAD boundaries do not completely constrain promoter interactions within TADs.

I then related the localisation of promoters and PIRs relative to TADs. This shows that promoters as well as PIRs tend to be located proximally to TAD boundaries. Unexpectedly, partner RFs tend to interact in a cross-TAD manner. I.e. when one RF is located proximally to a TAD boundary, its partner RFs tend to be located proximally to the opposite boundary. Moreover, if an RF is located centrally within a TAD, its partner RFs may be located proximally to either boundary at roughly equal frequencies. These findings show that promoter interactions form in a manner that is consistent with the loop extrusion model. I.e. the highest interaction frequencies are found at the base of the DNA loop where cohesin presumably physically links the ends.

To learn more about promoters with inter-TAD-only interactions, I examined the transcriptional activity of the corresponding genes and I found that these tend to be transcriptionally less active than those with intra-TAD-only interactions. However, differential transcript levels were absent when I excluded inter-TAD-only promoters

that were located proximally to a TAD boundary. This suggests that TAD boundary properties may convey transcriptional repression. Lastly, I investigated protein binding at PIRs and I found that PIRs of baits with inter-TAD-only interactions tend to be supported by TFs that possess transcription regulator activity. This suggests that these PIRs may be enhancers and that the inter-TAD-only promoter interactions may be involved in transcriptional regulation.

5.1.2 Dynamic TAD organisation may underpin apparent boundary crossing promoter interactions

In light of recent single-cell work on spatial genome organisation, it appears that TAD-like structures tend to form at genomic intervals that are marked by cohesin and CTCF, but may not be strictly limited to those regions (Cattoni *et al.*, 2017; Flyamer *et al.*, 2017; Bintu *et al.*, 2018). This suggests a much more flexible chromosomal organisation at the TAD level than previously presumed. Furthermore, this conceptually challenges the notion that TADs constitute a static genomic partitioning and rather, favours a dynamic landscape where by DNA interactions form with an increased frequency between loci mapping to TAD boundaries. I report that ~30% of promoter interactions cross TAD boundaries which certainly supports the notion that TAD boundaries insulate incompletely. However, perhaps at least some of the perceived TAD boundary crossing arises from conflating population signals, potentially as a result of detecting a promoter interaction that only exists in the absence of a TAD in a given cell. The fact that I still detect TAD boundary crossing promoter interactions in cell cycle synchronised cells (Chapter 3) is however in support of genuine TAD boundary, since much of the heterogeneity in spatial genome organisation likely arises from cell cycle effects (Nagano *et al.*, 2017). Altogether, it is likely that in the scope of TADs, chromosome organisation is dynamic and that population-derived TAD boundaries are likely loci involved in DNA looping, while other loci may form domains at atypical locations. Potentially, at the level of individual cells, promoters with inter-TAD-only interactions are exclusively contacted in absence of a typical TAD.

5.1.3 PIR features may influence TAD boundary crossing ability of promoter interactions

To further investigate inter-TAD-only promoter interactions, I evaluated protein binding and histone modifications at the PIRs, as well as the transcriptional activity of the corresponding genes. I show that genes whose promoters engage in inter-TAD

interactions with boundary-proximal PIRs show diminished transcriptional activity. This suggests a role for some TAD boundary-proximal loci in transcriptional repression. In part, this notion is supported by my observation that PIRs of cohesin dependent (lost) promoter interactions (which are mostly localised proximally to TAD boundaries) are enriched for the binding of PRC2 component EZH2 (Figure 50). Lastly, I found that the constitutive heterochromatin associated mark H3K9me3 is a significant predictor for PIRs of intra-TAD-only interactions (Figure 28). Since this mark is enriched in constitutively heterochromatic regions, this suggests that promoter interactions in constitutive heterochromatin are more restricted by TAD boundaries than those in other chromatin segments. My analyses did not implicate additional histone modifications in driving intra-TAD versus inter-TAD promoter interactions, however it must be noted that they were limited to eight modifications (H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9me3, H3K9ac, H3K27me3, and H3K79me3). Together, these findings beg the question: are inter-TAD promoter interactions involved in transcriptional repression? Perhaps the ability to cross TAD boundaries is related to the mechanism that stabilises these promoter interactions.

Interestingly, I found that the PIRs of inter-TAD interactions show DNA regulatory elements, suggesting that promoter-enhancer interactions may form across TAD boundaries. Conversely, I detect YY1 as a predictor of PIRs with intra-TAD interactions. This is in line with reports that YY1 connects looping interactions within TADs (Beagan *et al.*, 2017) and connects promoters and enhancers (Weintraub et al., 2017). Since YY1 has been implicated in connecting looping interactions in the context of developmental regulation (Beagan *et al.*, 2017; Weintraub *et al.*, 2017), it may be interesting to further investigate the target promoters of YY1 bound PIRs in a developmentally relevant system. However, as HeLa cells used in my thesis are incapable of differentiation, they are not an appropriate model for such analyses. Therefore, future experiments using a developmentally relevant model may shed light on the prevalence and persistence of inter-TAD promoter interactions, as well as their role in development. Such experiments should include analysis of transcriptional data to determine the effect of promoter interactions on transcriptional activity.

5.1.4 Cell cycle perspectives

In the analyses presented in Chapter 3 of this thesis, I focus mainly on promoter interactions that are invariant throughout the interphase stages G1 and G2. Although this approach improves confidence in detection of TAD boundary crossing events, it disregards some potentially interesting cell-cycle-specific promoter interactions and TAD boundary crossing events. For instance, the clustered interaction scores heat map (Figure 13) contains two clusters (C and D) showing G1- and G2-specific promoter interactions respectively. Although these clusters represent a minority of all clustered interactions (12.6% and 10.1% respectively), it may be interesting to evaluate them further. For instance, I find that promoter interaction distance increases from Mitosis (cluster E) to G1 and G2. This is consistent with my observations that TAD boundary crossing is more prevalent in G2 than G1 (Figure 14 and Figure 15) and that inter-TAD interactions tend to act over greater genomic distances (Figure Figure 27). However, this appears in conflict with previous Hi-C studies which have reported a decrease in DNA interaction distance throughout cell cycle progression (Dileep *et al.*, 2015; Wutz *et al.*, 2017). This discrepancy may be related to the difference between examining global pairwise DNA interactions of any locus versus promoter interactions, which may facilitate distal regulatory contacts. Indeed, it has been noted that transcription-compatible chromatin in G1 shows a higher probability of short-range contacts than long-range contacts (Barrington, Pezic and Hadjur, 2017) and that TAD boundaries in G2 insulate interactions to a lesser degree than in G1 (Nagano *et al.*, 2017).

In a separate but related observation, I find that promoter interactions may show differential TAD boundary crossing in G1 and G2 (Figure 20). Although the majority of inter-TAD-only promoter interactions show presence of a TAD boundary between the bait and PIRs in G1 as well as G2, large minorities of promoter interactions show presence of a TAD boundary between the bait and PIRs in G1 or G2 exclusively. This suggests that promoter interactions may persist in genomic regions where TAD organisation is dynamic, presumably as a result of genome duplication during S-phase.

Taken together, these observations leave exciting avenues for future analyses; do persistent promoter interactions ignore rearranging TADs? Is the observed promoter interaction distance discrepancy related to cell cycle progression and perhaps

supported by distinct proteins? A combined approach of single cell Hi-C and high temporal resolution bulk PCHi-C may help answer these questions by identification of persistent and dynamic promoter interactions and by allowing identification of genomic regions that likely harbour TADs, while simultaneously informing on atypical domain structures.

5.2 Promoter interactions in the context of architectural protein perturbations

5.2.1 Spatial properties of rewiring promoter interactions

In chapter 4.1 and 4.2, I show that upon cohesin depletion, different subsets of promoter interactions are *lost*, *maintained*, as well as *gained*. These rewiring categories show distinct promoter interaction distance profiles (Figure 36): *lost* promoter interactions typically act over longer genomic distances than *maintained* or *gained* promoter interactions. Additionally, with respect to TADs, frequency plots of categorised promoter interactions show distinct profiles (Figure 38): *lost* promoter interactions are typically found proximally to TAD boundaries, while *maintained* interactions are located anywhere within a TAD and *gained* interactions are predominantly located proximally to TAD boundaries while interacting across them. Furthermore, a comparatively small proportion of *maintained* promoter interactions were located within high-confidence TADs (Figure 42), implying that *maintained* interactions may be supported by spatial organisation other than TADs. These findings indicate that *lost* (cohesin dependent) promoter interactions are most strongly related to TAD organisation, reminiscent of “chromatin loops” (Rao *et al.*, 2014) and their response to cohesin depletion (Rao *et al.*, 2014). Interestingly, a recently published article provides evidence that cohesin- and CTCF-mediated distal looping interactions form the first step in creating interactions between promoters and “enhancer chains” (Song *et al.*, 2019). The cohesin- and CTCF-related nature of these looping interactions suggests that (sub-)TAD boundaries provide an initial stable connection between a promoter and an enhancer and subsequently to additional sets of enhancers that are connected via enhancer-enhancer interactions. Perhaps the importance of cohesin in the first step of setting up gene regulatory domains may be the reason why cohesin depletion leads to particularly prominent misregulation of inducible genes (Cuartero *et al.*, 2018) Furthermore, the selective

enrichment of cohesin and CTCF at the initial stable interacting elements suggests that promoter interactions with “enhancer chains” may be less reliant on cohesin *in cis*. In support of the notion that promoter-enhancer interactions may exist in the absence of cohesin, my results show two classes of promoter interactions that are likely cohesin-independent: *gained* and *maintained* interactions.

Gained promoter interactions tend to form across TAD boundaries, therefore they are likely ectopic promoter interactions that form in the absence of confinement by TAD boundaries consistent with previous reports (Nora *et al.*, 2012; Seitan *et al.*, 2013; Zuin *et al.*, 2014; Lupiáñez *et al.*, 2015; Flavahan *et al.*, 2016; Hanssen *et al.*, 2017; Rodríguez-Carballo *et al.*, 2017). *Maintained* promoter interactions appear unrestricted in terms of their location within TADs, although they do appear to be constrained by TAD boundaries (see Figure 38, middle row). This dissimilarity between *gained* and *maintained* promoter interactions raises the question: what could be the mechanisms underlying these two classes of cohesin-independent interactions?

One line of reasoning states that co-localisation of promoters and enhancers within the same TAD mainly drives promoter-enhancer interaction specificity (Dekker *et al.*, 2016). By extension, ectopic promoter interactions upon TAD boundary perturbations are expected to form primarily with active enhancers that are located beyond the TAD boundary. This is in line with my results, that show transcriptional up-regulation of genes with *gained* promoter interactions (Figure 53 and Figure 54), as well as characteristic features of enhancers at PIRs involved in *gained* interactions (inferred from the similarity between PIRs involved in *maintained* and *gained* promoter interactions; Figure 50). Furthermore, somatic mutations leading to a reduced capacity of TAD boundaries to prevent formation of ectopic promoter enhancer interactions has been linked to oncogenesis (Valton *et al.*, 2016), highlighting that the gain of functional ectopic interactions is directly relevant in non-academic “real-world” situations. Although these observations suggest that ectopic promoter interactions are by necessity functional (i.e. stimulating transcription), a recent study on the mouse ncRNA genes *Xist* and *Tsix* in mESCs provides evidence to the contrary (van Bemmel *et al.*, 2019). *Xist* and *Tsix* are are located across a TAD boundary and predominantly show DNA interactions within their cognate TADs. Van Bemmel et al. show that upon inversion of the genomic interval that houses these

genes, many of their wild type interactions are lost while novel interactions are gained within their new TADs. Surprisingly, Xist was transcriptional upregulated while Tsix was unaffected and remained actively transcribed. This highlights that the TAD environment may shape promoter interaction landscapes while not necessarily altering transcription. Since Tsix was transcriptionally unaffected by the inversion, its promoter appears insensitive to the *cis*-regulatory environment. Therefore, it appears to be possible to gain non-functional ectopic promoter interactions.

Interestingly, the wild type region that contains Tsix resembles a stripe on the Hi-C matrix (see: van Bemmel *et al*., 2019 Figure 1a). This stripe structure appears partially lost upon the inversion and de novo stripe formation appears absent (see: van Bemmel et al., 2019 Figure 1f and g). This type of genomic structure (Vian *et al.*, 2018) has been implicated in facilitating promoter-enhancer pairing, driving transcription in the context of development (Barrington *et al.*, 2019). Given that Tsix appears transcriptionally unaffected and Xist is transcriptionally upregulated, this suggests that the transcriptional activity of Tsix was not driven by stripe-supported promoter interactions whereas the upregulation of Xist may have been supported by the (remnants of) the stripe. Stripes are likely formed by asymmetric loading of cohesin and are related to CTCF binding and the cohesin loading factor NIPBL (Vian *et al.*, 2018; Barrington *et al.*, 2019). Barrington et al. suggest that stripes may represent the population Hi-C readout of a process called tracking (Blackwood *et al.*, 1998). In this mechanism, activator proteins bind to an enhancer and slide along the DNA while maintaining bound to the enhancer element, thereby forming a loop (similar to loop extrusion). Once a promoter is encountered, tracking halts and the promoter is activated. Tracking offers an exciting possibility for the formation of ectopic promoter interactions in the absence of cohesin. Indeed, if tracking is limited to neighbouring promoters and enhancers, this might explain the limited number of *gained* promoter interactions that I detect. However, so far cohesin has been proposed as the most likely tracking factor, therefore for this model to hold true, *gained* promoter interactions would require an alternative, yet unidentified tracking mechanism.

Additionally, the matter of cohesin depletion resistant (*maintained*) promoter interactions adhering to TAD intervals remains an open question. One possibility is that *maintained* promoter interactions are formed in a manner that requires cohesin/

CTCF and are stabilised once established, alleviating their direct reliance on these proteins. It has been suggested that cohesin-depletion-resistant DNA loops are supported by interactions with superenhancers (Rao *et al.*, 2017), however, the interaction profile of *maintained* promoter interactions does not resemble loops, nor interactions with clusters of enhancers (see Figure 38). Furthermore, the linear genomic distance over which the loops reported by Rao et al. act is three orders of magnitude larger than the *maintained* promoter interactions that I detect (23.15Mb versus 56kbp). Together, this suggests cohesin-depletion-resistant loops are not supported by the same mechanism as *maintained* promoter interactions. An attractive alternative explanation is offered by “enhancer chains”, presuming that they may persist upon removal of the initial guiding enhancer interaction. However, the stability of *maintained* promoter interactions is unknown. It is possible that I detected these interactions by virtue of the rapid action of the Degron system (30 minutes), implying that *maintained* promoter interactions may disappear upon longer incubation with Auxin. However, if a prolonged absence of cohesin results in greater loss of promoter interactions, I expect this to be reflected in the number of genes that are misregulated (specifically, downregulated). The SLAM-Seq results presented in chapter 4.4 show a total of 687 misregulated genes (421 up and 266 down). The largest number of misregulated genes upon cohesin perturbation in the literature (1984; 959 up, 1025 down) was reported after Rad21 RNAi knockdown for a duration of eight days (Schaaf *et al.*, 2009). Although eight days easily qualifies as a prolonged period, it may be incorrect to compare this experiment to my results since it was performed in *D. melanogaster*. Additionally, RNAi knockdown may incompletely remove functional cohesin. In a separate report, rapid depletion of the cohesin subunit RAD21 led to transcriptional misregulation of 1461 genes (534 up and 927 down) at a $\log_2$ fold change cut-off of 1.3 (and $p < 0.05$) after 6 hours (Rao *et al.*, 2017), whereas I used a $\log_2$ fold change cut-off of 0.1 (and FDR ≤ 0.05) for the SLAM-seq data. Although such discrepancies may be attributed to the used analysis techniques, they are unlikely related to the method of cohesin perturbation, as Rao et al. also depleted cohesin by using an AID system. Furthermore, Rao et al. used PRO-Seq, which also measures nascent transcript levels. However, PRO-Seq measures RNA that is transcribed by Pol I, II and III, in contrast to SLAM-seq which measures RNA that is transcribed by Pol II only (Wissink *et al.*, 2019). This may have contributed to the larger number of differentially transcribed genes reported by

Rao et al. A feasible explanation for the observed fold-change discrepancies is that the SLAM-seq 4-thiouridine nucleotide conversion efficiency was low, leading to a reduced nascent transcription signal. Yet, even the larger extent of downregulation reported by Schaaf et al. and Rao et al. is limited to a small proportion of all genes (~13% in drosophila versus ~7% in humans). Furthermore, prolonged exposure to absence of cohesin may result in secondary regulatory effects, for example through cellular stress response. This may result in transcriptional differences that are not the direct result of DNA interaction rewiring. Additionally, direct comparison of the reported numbers of misregulated genes is problematic due to experimental differences. E.g. statistical definitions of differentially regulated genes were not the same and different techniques were used to measure transcript levels. Collectively, it appears that many *maintained* promoter interactions can persist upon prolonged absence of cohesin. One exciting prospect is the identification of factors that may support such interactions. It will be interesting to see if perturbation of such factors (in absence of cohesin) would result in loss of *maintained* as well as *gained* promoter interactions. Further research into ectopic cohesin-independent (*gained*) promoter interactions may shed light on factors involved in their formation, possibly supported by the tracking mechanism. One potential protein is the bivalent TF YY1, which I detected as a significant predictor of inter-TAD-only promoter interactions (see Figure 27; note that this is in the presence of cohesin). However, although YY1 has been implicated in facilitating spatial DNA organisation (Beagan *et al.*, 2017) and promoter-enhancer loops (Weintraub *et al.*, 2017), it seems that YY1 may operate similarly to CTCF, i.e. by halting cohesin-based loop extrusion. This suggests that the role of YY1 in shaping DNA interactions may depend on an extruding/tracking factor such as cohesin. Another possibility is that the mediator complex is involved in supporting/maintaining promoter interactions in absence of cohesin. Since the mediator complex transmits regulatory signals from enhancer-bound TFs to PICs bound at promoters it has been widely proposed as a factor in DNA loop formation (Malik *et al.*, 2010; Allen *et al.*, 2015; Soutourina, 2018). However, a recent investigation provides evidence that the mediator complex may relay transcriptional signals without physically tethering DNA elements, presumably relying on cohesin to bring promoters and enhancers within sufficient range for TFs to transmit regulatory signals to the mediator complex by diffusion (El Khattabi *et al.*, 2019). Furthermore, El Khattabi et al. find that depleting RAD21 and the Pol II subunit RPB1

simultaneously, does not result in greater loss of promoter-enhancer interactions than upon RAD21 depletion alone, showing that RPB1 not involved in facilitating cohesin-independent promoter interactions. Yet another possibility is that the cohesin paralogs condensin I or condensin II may support promoter interactions by tethering them, similarly to- but in the absence of cohesin. This provides an attractive mechanism, since condensins are known to be capable of binding to DNA and have even been shown to actively extrude DNA in an ATP-dependent fashion in vitro (Ganji *et al.*, 2018). Although condensin I was thought to be sequestered in the cytoplasm during interphase (Hirota *et al.*, 2004), more recent work has been able to detect nuclear condensin I in interphase by ChIP-Seq (Li *et al.*, 2015) and subsequently by microscopy (Zhang *et al.*, 2016). While this supports condensin I as a potential candidate for supporting *maintained* and *gained* promoter interactions in the absence of cohesin, condensin II is more abundant in the interphase nucleus and has been detected at promoters as well as enhancers (Dowen *et al.*, 2013), and therefore condensin II is potentially the more promising factor. However, a recent preprint shows that perturbation of condensin II does not lead to changes in spatial genome organisation and shows limited transcriptional changes (Abdennur *et al.*, 2018). Although this suggests no involvement in transcriptional regulation as well as spatial genome organisation, Abdennur et al. do not evaluate the simultaneous perturbation of cohesin and condensin II. Since it is possible that cohesin may preserve spatial organisation upon condensin II perturbation, their results do not exclude condensin II in the context of supporting *maintained* and *gained* promoter interactions. In conclusion, more research still has to be done to investigate factors involved in supporting cohesin independent promoter interactions. The sets of *maintained* and *gained* promoter interactions that I describe in this thesis offer an exciting starting point in the search for potential proteins and mechanisms.

5.2.2 Promoter interaction rewiring and chromatin features

In chapter 4.3 I analysed ChIP-Seq data to explore which factors are indicative of promoter interaction loss, maintenance or gain upon cohesin depletion. I used two approaches for relating ChIP-Seq targets to promoter interaction rewiring: LASSO logistic regression and a Fisher-test-based overrepresentation analysis based on thresholded ChIP signals (“95th percentile analysis”). The LASSO logistic regression analyses are likely more stringent than the 95th percentile analyses, owing to the

property of LASSO regression to mitigate multicolinearity and to control for promoter interaction distance and score (see Figure 48 and Figure 49). Therefore, I will focus on the LASSO results in this discussion. However, it should be noted that the 95th percentile analysis provided two insightful results: (1) Baits and PIRs show similar patterns of ChIP-Seq signal overrepresentation (see Figure 50). This indicates that promoters and distal elements in these analyses show broadly similar properties in terms of protein binding and histone modifications. Interestingly, H3K4me1 appears only at PIRs in the 95th percentile analysis, highlighting the enhancer-specificity of this mark. (2) RFs involved in *maintained* and *gained* promoter interactions show similar patterns of ChIP-Seq signal overrepresentation. This indicates that *maintained* and *gained* promoter interactions are similar in terms of protein binding and histone modifications at the interaction partners, suggesting that these two types of promoter interactions may be supported by similar mechanisms. In other words, a *gained* promoter interaction may be a *maintained*-like promoter interaction that is no longer restrained by a TAD boundary.

In the LASSO regression analyses I decided to focus on *maintained* versus *lost* promoter interactions for two reasons: because there are more *maintained* than gained promoter interactions (which provides greater statistical power in the regression analysis), and because *maintained* promoter interactions provide an exciting opportunity to study transcriptional robustness to cohesin depletion. It should be noted that an alternative solution would have been to use LASSO multinomial regression. However, I chose against this option because the promoter interaction rewiring dataset is strongly imbalanced in terms of the numbers of *gained* versus *lost* and *maintained* interactions. I decided to perform separate LASSO regression analyses on histones/histone modifications and TFs/DNA-associated proteins because I wanted separate feature selection procedures to be applied to histones/histone modifications and DNA binding proteins. The LASSO logistic regression analyses revealed sixteen proteins and five histones/histone modifications as significant predictors for *maintained* versus *lost* promoter interactions. Since Gene Ontology analyses provided limited results (see chapter 4.3), I searched the literature with the aim of identifying proteins among the significant predictors of either interaction class that have known roles in

transcriptional regulation and, in the case of *maintained* promoter interactions, in potentially supporting cohesin-independent interactions.

5.2.2.1 Proteins involved in cohesin-dependent promoter interactions

Firstly, since RFs involved in *lost* promoter interactions most strongly conform to TAD organisation, it is encouraging to find CTCF and Cohesin ChIP-Seq signal at PIRs to be predictive of this category of promoter interaction. Additionally, I find that the histone variant H2AFZ (also referred to as H2A.Z.1) relates to PIRs of *lost* promoter interactions. Since H2AFZ is strongly enriched at nucleosomes surrounding CTCF binding sites (Millau *et al.*, 2011), this supports the notion that *lost* promoter interactions are related to cohesin/CTCF-mediated loops and TADs. In terms of transcriptional regulation, PIRs that are involved in *lost* promoter interactions may support transcriptional activity through GA-binding protein (GABP) and host cell factor (HCFC1). GABP is a known transcriptional activator of housekeeping- as well as lineage-specific genes (Rosmarin *et al.*, 2004), it is necessary for cell cycle progression in the interphase (Yang *et al.*, 2007), and it interacts with HCFC1 (Rosmarin *et al.*, 2004). Note that my results are based on the GABP subunit alpha; GABPA. HCFC1 is a transcriptional coactivator that is also involved in cell cycle progression and that is known to indirectly localise at promoters by interacting with other TFs such as GABP and YY1 (Yu *et al.*, 2010; Michaud *et al.*, 2013). The fact that GABP and HCFC1 ChIP-Seq signal at PIRs is predictive of *lost* promoter interactions suggests that these proteins may stimulate transcriptional activity through promoter interactions at TAD boundaries and supports the notion that TADs may be involved in DNA replication timing (Pope *et al.*, 2014). However, in a recent paper from the same lab it was shown that TAD boundaries are dispensable for DNA replication timing, while pointing out that CTCF-independent intra-TAD DNA interactions with distinct “early replication control elements” are responsible for driving DNA replication timing while simultaneously being involved in shaping TAD organisation in a CTCF-independent manner (Sima *et al.*, 2019). Lastly, I find that RPA1 and RPA2 ChIP-Seq signal is predictive of PIRs that are involved in *lost* promoter interactions. RPA1 and 2 are historically known to be involved in DNA replication and repair (Wold, 1997), although additional investigations have suggested a direct interaction between RPA proteins and Pol II (Sikorski *et al.*, 2011)

and have implicated RPA1 and RPA2 in promoter decompaction in the context of transcriptional regulation (Fujimoto *et al.*, 2012). Furthermore, RPA1 and RPA2 have recently been suggested to indirectly recruit histone H3.3 to enhancers and promoters in the context of transcriptional regulation (Zhang *et al.*, 2017). Collectively, the proteins that I have found to be predictive of *lost* promoter interactions appear to be related to TAD organisation, transcriptional regulation and cell cycle progression. Assuming that *lost* promoter interactions are directly related to TADs, the proteins GABP, HCFC1, RPA1 and RPA2 offer an exciting view on the mechanisms by which TADs may regulate transcriptional activity. Additionally, these proteins may be related to transcriptional regulation *in cis* and, assuming that they depend on cohesin to localise to PIRs of *lost* promoter interactions, they may be related to the observed changes in chromatin accessibility at PIRs upon cohesin depletion.

5.2.2.2 PIRs involved in maintained promoter interactions Polycomb and NuRD

The presence of EZH2 and MBD3 in the set of significant predictors for PIRs of *maintained* promoter interactions was striking, since they are components of two well-known repressive complexes: the PRC2 component EZH2 and the NuRD component MBD3 (Denslow *et al.*, 2007; Margueron *et al.*, 2011). NuRD has been shown to indirectly recruit PRC2 through H3K27 deacetylation (Reynolds *et al.*, 2012), indicating that these two repressive complexes may perform their functions cooperatively. Furthermore, I found that the NuRD-related protein PWWP2A is also predictive of PIRs with *maintained* promoter interactions. However, the latter occurs predominantly at active promoters (Pünzeler *et al.*, 2017). PWWP2A has recently been shown to compete with MBD3 for association with the NuRD complex and, in that capacity, may promote transcription elongation by Pol II (Zhang *et al.*, 2018). Therefore, the detection of both MBD3 and PWWP2A in the context of *maintained* promoter interactions suggests a role for NuRD-related transcriptional fine-tuning. The Polycomb proteins, on the other hand, are known to be involved in nuclear architecture (Tiwari *et al.*, 2008; Bantignies *et al.*, 2011) and may be directly involved in tethering DNA interactions: a recent preprint reports that the PRC directly supports long-range DNA interactions in the absence of cohesin in human embryonic stem cells, suggesting that cohesin works antagonistically to the PRC by preventing it from forming these long-range DNA interactions (Rhodes *et al.*, 2019). Whether PRC is

also directly involved in supporting some short-range *maintained* promoter interactions is an exciting question for future analyses. My finding that PIRs bearing poised/repressed annotations are involved in shorter range *maintained* promoter interactions than those bearing active annotations suggests a role for repressive factors (see Figure 37). Involvement in transcriptional processes appears as a common property of many of the features related to PIRs with *maintained* promoter interactions. Most prominently: histone acetyltransferase P300 and bromodomain-containing protein 4 (BRD4).

5.2.2.2.1 P300

P300 is involved in a plethora of mechanisms and processes, most notably: cell cycle progression, proliferation, differentiation, apoptosis, tumorigenesis and transcriptional regulation (Goodman *et al.*, 2000; Ho Man Chan *et al.*, 2001). P300 is canonically considered an enhancer mark (Visel, Blow, *et al.*, 2009), although it is found at promoters as well as enhancers (Heintzman *et al.*, 2007) and it appears to be dependent on other proteins to be able to stably bind DNA, e.g. p53 (Lill *et al.*, 1997). It is known to interact with a wide range of other proteins, including core-transcriptional machinery such as the PIC component TBP (Ho Man Chan *et al.*, 2001), which is among the significant predictors of PIRs with *maintained* promoter interactions. P300 has been suggested to form a scaffold for transcriptional signalling (Ho Man Chan *et al.*, 2001). In addition, P300 has histone acetyltransferase (HAT) activity, which is thought to promote DNA accessibility for TFs (Ho Man Chan *et al.*, 2001). The active mark H3K27ac can be deposited at enhancers by P300 (Creyghton *et al.*, 2010), which is in line with my results (compare Figure 48 and Figure 49). Although P300 is generally known as a transcriptional coactivator, it may also suppress transcriptional activity. For instance, P300 can deposit SUMO, which leads to transcriptional downregulation (Girdwood *et al.*, 2003). Additionally, P300 and the histone deacetylase HDAC3 have been shown to cooperatively downregulate transcription of the oncogene c-Myc (Sankar *et al.*, 2008). Taken together, the known roles of P300 in transcriptional regulation suggest that the PIRs of *maintained* promoter interactions show enhancer characteristics.

5.2.2.2.2 BRD4

BRD4 is a protein with similarly multifaceted roles. BRD4 binds chromatin throughout the cell cycle (Dey *et al.*, 2003) and it locates to acetylated histone H4

lysine residues 5,8 and 12 (Hargreaves *et al.*, 2009). Additionally, BRD4 can bind to other proteins through its N-terminus, forming a molecular scaffold (Wang *et al.*, 2012). BRD4 is involved in transcriptional regulation of primary response genes (Hargreaves *et al.*, 2009), presumably through direct phosphorylation of the Pol II CTD (Devaiah *et al.*, 2012) or through indirect phosphorylation of the Pol II CTD by the elongation factor pTEFb (Patel *et al.*, 2013). Interestingly, P300 may be involved in the latter process through acetylation of RelA, which activates BRD4-pTEFb whereupon Pol II is phosphorylated (Huang *et al.*, 2009), suggesting that BRD4 and P300 may cooperate to regulate transcription. Furthermore, BRD4 is involved in chromatin organisation through its capacity to evict nucleosomes leading to decompaction (Devaiah *et al.*, 2016) and through its role as a chromatin insulator, limiting the spread of the histone variant γH2AX by recruiting the condensin II complex in the context of the DNA damage response (Floyd *et al.*, 2013). The role of BRD4 in chromatin organisation is further highlighted by reports of widespread chromatin decompaction upon BRD4 perturbation (Wang *et al.*, 2012; Floyd *et al.*, 2013). These results suggest a role for BRD4 and condensin II in supporting cohesin independent DNA interactions. Although the BRD4 paralog BRD2 has been shown to cooperate with CTCF in the context of spatial genome organisation, BRD4 is specifically reported to not co-localise with CTCF (Hsu *et al.*, 2017). This further suggests that BRD4 and potentially condensin II recruitment, may function in a cohesin-independent manner. Collectively, the relationship between BRD4 and *maintained* promoter interactions suggests that the PIRs involved in these interactions are related to transcriptional regulation. Additionally, the capacity of BRD4 to recruit condensin II suggests a role for the latter in supporting cohesin-independent promoter interactions. However, as mentioned above, a recent preprint has shown that condensin II depletion does not affect TADs and that transcription of only a limited number of genes is affected (Abdennur *et al.*, 2018). The lack of a double perturbation of cohesin and condensin II leaves the possibility that the contribution of condensin II to spatial chromatin organisation is largely redundant and that upon its depletion, spatial organisation is preserved by cohesin.

Taken together, these results show that PIRs that are involved in *maintained* promoter interactions appear to associate with transcriptionally relevant proteins.

This is further supported by the presence of the canonical enhancer histone marks H3K4me1 and H3K27ac in the LASSO results (see Figure 49). However, I also detect promoter-associated features as significant predictors for the PIRs of *maintained* promoter interactions. For instance, H3K4me3 is widely considered to be a mark of promoter regions. Additionally, the TATA-box binding protein TBP is a core part of the TFIID component of the PIC (Thomas *et al.*, 2006) and the DNA helicase component CHD2 is known to recruit to promoters in mice (Siggens *et al.*, 2015). However, in recent years, it has become increasingly evident that the distinction between promoters and enhancers is not clear cut; H3K4me3 may be found at distal elements, and enhancers may be transcribed by Pol II to produce eRNAs (Andersson, 2015). In line with this notion, the transcriptional activators GATAD1 and PHF8 (which I find to be predictive of *maintained* promoter interactions) have been detected at promoter regions as well as non-promoter regions marked by H3K4me3 (Vermeulen *et al.*, 2010). Collectively, the detection of these factors in association with PIRs of *maintained* promoter interactions suggests that these PIRs likely contain *active* enhancers that associate with transcriptional machinery. Given that our PCHiC data suggest that these PIRs are located proximally to promoters through looping interactions, this may be indicative of transcriptional hubs. Investigating the mechanism that establishes or supports *maintained* promoter interactions in the absence of cohesin is an extremely interesting and relevant question that needs to be addressed in future research.

5.2.3 Transcriptional response and differential chromatin accessibility upon cohesin depletion

In chapter 4.4, I asked how the transcriptional response as well as differential chromatin accessibility upon cohesin depletion relate to promoter interaction rewiring. To this end, I assessed the dynamics of nascent transcription profiled by SLAM-seq (Herzog *et al.*, 2017), as well as differential chromatin accessibility by ATAC-seq (Buenrostro *et al.*, 2013). The combination of these two methods is particularly powerful in the context of rapid cohesin depletion, owing to the detection of novel transcripts by SLAM-seq (thereby excluding the signal from retained, steady-state RNA) and to the precise base-pair resolution of TN5 integration in ATAC-Seq. With this approach, I found that downregulated genes upon cohesin depletion show reduced chromatin accessibility at their baits, and upregulated genes upon cohesin depletion show increased chromatin accessibility at their baits (Figure

58). Furthermore, I found that the combined effects of promoter interaction rewiring and differential compaction at PIRs can explain the transcriptional response (Figure 59). Notably, I find that chromatin accessibility at promoters and PIRs appears to adapt to cohesin depletion. This suggests that chromatin accessibility may be shaped by cohesin. This is interesting, as it complements the observations that place architectural protein binding downstream of the chromatin state. For example, it has recently been shown that deposition of repressive marks H3K9me3 and DNA methylation (which are related to condensed chromatin) can evict CTCF (Tarjan *et al.*, 2019) and that H3K4me1 deposition may recruit cohesin (Yan *et al.*, 2018). However, these findings are not necessarily contradictory with my data. For example, if promoter-enhancer interactions provide sufficient proximity between DNA elements to enable an attached protein-complex to spread its chromatin-modifying effects onto the interacting elements, loss of an interaction may result in a loss of chromatin modification at an interacting partner. It is an interesting prospect to investigate whether epigenetic signals may be communicated from promoters to enhancers in a feedback loop, resulting in the differential accessibility at PIRs in my results. Perhaps activation of enhancers by promoters may relate to “enhancer chains” (Song *et al.*, 2019). Such a mechanism may be beneficial, for instance, for sequential activation of genes. I.e. if an inactive promoter interacts with a chained enhancer which is activated by another promoter, this may communicate transcriptional activity from one gene to another. One approach to investigate this is to perform anti-H3K27ac ChIP-Seq, in addition to enhancer CHiC (where baits are used to enrich a Hi-C library for interactions with enhancers) in time-course in a differentiating cell model. Identification of sequential deposition of the H3K27ac mark at enhancers and interacting promoters may shed light on the direction in which activating signals may be transmitted. To establish whether these signals result in functional regulatory signal transmission, this approach can be complemented by nascent RNA sequencing.

5.2.4 Perspectives on promoter interactions from mitotic chromatin

In chapter 4.1, I use a combined approach of k-means clustering and subsequent differential interaction calling (using Chicdiff) to arrive at a high-confidence set of promoter interactions that are *lost*, *maintained* or *gained* upon depletion of the

cohesin subunit SCC1. I complement these results with promoter interaction data from mitotic cells and WAPL/PDS5A/PDS5B RNAi conditions.

It would be interesting to analyse mitotic promoter interaction data in more detail in future work. For instance, there are indications that during mitosis, transcriptional activity is not entirely absent (Palozola *et al.*, 2017) and that promoters may retain accessibility (Hsiung *et al.*, 2015). Although it has been shown using Hi-C that long-range DNA interactions are largely lost in mitosis (Naumova *et al.*, 2013), I do detect specific promoter interactions with PCHiC (igure 29, clusters E, J, M), including those detected exclusively in mitosis (cluster M). The possibility of mitotic transcription being supported by distinct promoter interactions is currently unexplored and may offer new insights into cell cycle processes and transcriptional regulation in general. It would also be interesting to compare mitotic promoter interactions to those that form in the semi-condensed “vermicelli” phenotype that is the result of WAPL/PDS5A/PDS5B RNAi (Tedeschi *et al.*, 2013). Specifically it may be of interest to relate the WAPL/PDS5A/PDS5B RNAi-specific promoter interactions (cluster I) to known transcriptional effects of WAPL perturbation (Tedeschi *et al.*, 2013). Collectively, the detected promoter interactions in mitotic chromatin offer an exciting avenue of further investigation, which was unfortunately out scope in this thesis.

6 Future directions

In this thesis I used a genome-wide approach to analysing the interplay between promoter interactions and architectural protein-dependent higher-order chromatin organisation. In chapter 3, I analysed promoter interactions with respect to TADs, identifying that promoter interactions tend to be located proximally to TAD boundaries and revealing a class of promoters that interact exclusively with DNA fragments spanning their cognate TAD boundaries. In chapter 4 I analysed promoter interaction rewiring upon rapid depletion of cohesin subunit SCC1. I identified promoter interactions that are *lost*, *maintained*, and *gained* under these conditions, and show that they have different properties with respect to linear interaction distance and chromatin features at PIRs. In particular, I demonstrate that the PIRs of cohesin independent promoter interactions have features of active enhancers and identify a set of candidate proteins that may be involved in supporting promoter-

enhancer interactions in the absence of cohesin. As such, these proteins provide a valuable starting point for the discovery of mechanisms underlying cohesin-independent interactions. Furthermore, I show that the immediate transcriptional response to cohesin depletion can be explained by the rewiring of interactions between promoters and active enhancers, and that chromatin accessibility at both promoters and PIRs is altered upon cohesin depletion. Reduced chromatin accessibility at enhancers losing promoter interactions in the absence of cohesin suggests that either cohesin contributes to enhancer activation *in cis,* or that promoters can communicate activating signals “back” to enhancers through cohesin-dependent chromosomal contacts. Finally, I describe the ongoing validation experiments confirming a functional role of cohesin-dependent promoter interactions in transcriptional control.

To expand on these findings, in future work it will be interesting to perform analyses in the context of inducible gene expression, such as upon immune cell activation or differentiation. While role of cohesin in inflammatory gene activation has been recently demonstrated (Cuartero *et al.,* 2018), the role of cohesin-independent promoter interactions in transcriptional control, remains unexplored. This includes investigations into the possible role of condensin in supporting such promoter interactions, particularly by means of “tracking” (Blackwood and Kadonaga, 1998) or formation of enhancer chains (Song *et al.,* 2019).

6.1 Beyond HeLa cells as the experimental model

The results in this thesis primarily derive from analyses that were performed in HeLa cells. This cell line provides a practical model for analyses in molecular biology, owing to its strong proliferative capacity and its amenability to genetic manipulation and RNAi. However, the HeLa cell line has serious limitations as a biological system. First, HeLa cells show karyotypic abnormalities such as extensive aneuploidy, chromosomal translocations and copy number aberrations (CNAs) (Adey *et al.*, 2013; Landry *et al.*, 2013). Theoretically, aneuploidy and chromosomal translocations are not a major problem for Hi-C and PCHiC analyses, since the former will be limited to systematic biases in sequencing depth (between chromosomes with differing ploidy), while the latter will lead to increased proportions of detected trans-chromosomal DNA interactions. This is expected to lead to the overall increased stringency of signal detection in PCHiC by CHiCAGO, since this

method uses the proportion of trans- to cis-chromosomal interactions to estimate technical noise. Therefore, the Poisson component of the background model is expected to be inflated. In addition, many trans-chromosomal promoter interactions detected using the reference genome reflect HeLa-specific artefacts, and therefore these interactions have been excluded from analysis after interaction calling with CHiCAGO. Additionally, it is noteworthy to draw attention to the fact that the approach to calculating ChIP-Seq scores per RF includes a normalisation step that corrects for systematic differences in sequencing depth (as a result of aneuploidy) (see Methods 2.11). CNAs do present a problem when aligning sequencing data against the human reference genome (here: GRCh37), as highly duplicated regions may result in artificially inflated read counts. However, promisingly, Hi-C and 5C have been used successfully to study genome organisation in HeLa, showing that the karyotype is not entirely prohibitive for C-based analysis strategies (Naumova *et al.*, 2013; Rao *et al.*, 2014; Wutz *et al.*, 2017). Nonetheless, using the HeLa-specific reference genome may resolve at least some of the artefacts. Due to the privacy-sensitive nature of the HeLa cell line, this reference genome has been restricted from public access. We have eventually procured access rights, however, by the time they were granted, the majority of my analyses had already been performed using GRCh37. In future work, I will use the HeLa reference genome to discard RFs that map to regions with CNAs and highly rearranged loci.

A second important limitation of the HeLa model for our analysis is that they do not represent an inducible system, whereas the role of cohesin in gene expression was shown to be particularly prominent for inducible genes (Cuartero et al., 2018). That study used mouse monocyte-to-macrophage differentiation, which provides an excellent model for such analyses (Cuartero *et al.*, 2018). It will be interesting to also study the role of cohesin-independent promoter interactions in this system. This may shed light on the role of cohesin-independent promoter interactions in initiating and maintaining transcriptional programmes. It will also be interesting to analyse the order, in which cohesin-dependent and independent enhancers connect to promoters, particularly given recent evidence suggesting that architectural proteins may stabilise promoter interactions with the first enhancer in the “enhancer chains” jointly transmitting regulatory signals to promoters (Song et al., 2019).

6.2 Possible mechanisms that support cohesin independent promoter interactions

My results suggest three promising possibilities for mechanisms that potentially support cohesin-independent promoter interactions.

First, condensin II is a likely candidate because it can tether DNA interactions and actively extrude loops (Ganji *et al.*, 2018). Furthermore, condensin II is recruited by BRD4 (Floyd *et al.*, 2013), which I find to be a significant predictor for *maintained* promoter interactions. To do this, it will be necessary to deplete cohesin and condensin II simultaneously – ideally by targeting AID to subunits of both complexes. Combined with an inducible vector system that can selectively rescue synthesis of the depleted subunit by introducing transcripts that do not encode the degron, selective rapid depletion of cohesin or condensin is made possible.

Secondly, transcriptional machinery itself may be responsible for supporting cohesin-independent promoter interactions. In support of this notion, inhibition of Pol II c-terminal domain phosphorylation by 5,6-dichlorobenzimadazole riboside (DRB) was reported to block “tracking” between the prostate-specific antigen receptor gene and a neighbouring enhancer, while simultaneously showing a reduction in TF and cofactor binding to the enhancer, including P300 (Wang *et al.*, 2005). Potent and rapid depletion of Pol II can be achieved with triptolide (Vispé *et al.*, 2009), which leads to proteasome-mediated degradation of the largest Pol II subunit RPB1 (Wang et al., 2011). Since I find the Pol II binding to be a significant predictor of *maintained* promoter interactions, we are currently exploring the effect of triptolide treatment on promoter interactions.

Finally, the Mediator complex remains a candidate for supporting cohesin independent promoter interactions despite the recent indications to the contrary (El Khattabi *et al.*, 2019). Notably, El Khattabi et al. do not specifically deplete Mediator but disrupt the mediator-Pol II complex by targeting Pol II with α-amanitin. In a similar approach to the double depletion strategy described above, a double-AID system that targets cohesin and one or more of the mediator subunits may successfully perturb mediator in addition to cohesin.

Altogether, in recent years a lot of exciting new discoveries have been made in the field of spatial chromatin organisation and its implications for transcriptional

regulation. The rapid development of new techniques and the refinement of existing ones has created a wealth of data and new analysis strategies. For instance, the emerging field of single cell Hi-C (Nagano *et al.*, 2013; Flyamer *et al.*, 2017; Ramani *et al.*, 2017) and super-resolution microscopy (Bintu *et al.*, 2018; Maass *et al.*, 2018) are providing invaluable insights that could not have been solved with bulk analyses. Although the work presented in this thesis relies on bulk analyses, my findings shed light on the principles that govern promoter interactions that can be further explored using these emerging methods.

7 References

Abdennur, N., Schwarzer, W., Pekowska, A., Shaltiel, I. A., Huber, W., Haering, C. H., Mirny, L. and Spitz, F. (2018) 'Condensin II inactivation in interphase does not affect chromatin folding or gene expression', *bioRxiv*, p. 437459. doi: 10.1101/437459.

Adelman, K. and Lis, J. T. (2012) 'Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.', *Nature reviews. Genetics*, 13(10), pp. 720–731. doi: 10.1038/nrg3293.

Adey, A., Burton, J. N., Kitzman, J. O., Hiatt, J. B., Lewis, A. P., Martin, B. K., Qiu, R., Lee, C. and Shendure, J. (2013) 'The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line', *Nature*, 500(7461), pp. 207–211. doi: 10.1038/nature12064.

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T. and Thanos, D. (2000) 'Ordered recruitment of chromatin modifying and general transcription factors to the IFN-β promoter', *Cell*, 103(4), pp. 667–678. doi: 10.1016/S0092-8674(00)00169-0.

Alipour, E. and Marko, J. F. (2012) 'Self-organization of domain structures by DNA-loop-extruding enzymes', *Nucleic Acids Research*, 40(22), pp. 11202–11212. doi: 10.1093/nar/gks925.

Allen, B. L. and Taatjes, D. J. (2015) 'The Mediator complex: A central integrator of transcription', *Nature Reviews Molecular Cell Biology*, 16(3), pp. 155–166. doi: 10.1038/nrm3951.

Andersson, R. (2015) 'Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model', *BioEssays*, 37(3), pp. 314–323. doi: 10.1002/bies.201400162.

Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) 'Telomeric Repeat–Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends', *Science*, 318(5851), pp. 798–801. doi: 10.1126/science.1147182.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a β-globin gene is enhanced by remote SV40 DNA sequences', *Cell*, 27(2), pp. 299–308. doi: 10.1016/0092-8674(81)90413-X.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*, 21(3), pp. 381–395. doi: 10.1038/cr.2011.22.

Bantignies, F. and Cavalli, G. (2011) 'Polycomb group proteins: Repression in 3D', *Trends in Genetics*, 27(11), pp. 454–464. doi: 10.1016/j.tig.2011.06.008.

Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J. and Hadjur, S. (2019) 'Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology', *Nature Communications*, 10(1), pp. 1–14. doi: 10.1038/s41467-019-10725-9.

Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V, Gillis, D. R. and Phillips-cremins, J. E. (2017) 'YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment', *Genome*

Research, 27(7), pp. 1139–1152. doi: 10.1101/gr.215160.116.Freely.

Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., Chang, Y., Li, J. B., Senaratne, T. N., Williams, B. R., Rouillard, J. M. and Wu, C. T. (2012) 'Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes', *Proceedings of the National Academy of Sciences of the United States of America*, 109(52), pp. 21301–21306. doi: 10.1073/pnas.1213818110.

Bell, A. C., West, A. G. and Felsenfeld, G. (1999) 'The protein CTCF is required for the enhancer blocking activity of vertebrate insulators', *Cell*, 98(3), pp. 387–396. doi: 10.1016/S0092-8674(00)81967-4.

Bengtsson, H., Jönsson, G. and Vallon-Christersson, J. (2004) 'Calibration and assessment of channel-specific biases in microarray data with extended dynamical range', *BMC Bioinformatics*, 5(1), p. 177. doi: 10.1186/1471-2105-5-177.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C. T., McBride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M. M., *et al.* (2009) 'Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence', *Nature Genetics*, 41(3), pp. 359–364. doi: 10.1038/ng.329.

Benko, S., Gordon, C. T., Mallet, D., Sreenivasan, R., Thauvin-Robinet, C., Brendehaug, A., Thomas, S., Bruland, O., David, M., Nicolino, M., Labalme, A., Sanlaville, D., Callier, P., Malan, V., Huet, F., Molven, A., Dijoud, F., Munnich, A., Faivre, L., *et al.* (2011) 'Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development', *Journal of Medical Genetics*, 48(12), pp. 825–830. doi: 10.1136/jmedgenet-2011-100255.

Berger, S. L. (2007) 'The complex language of chromatin regulation during transcription', *Nature*, 447(7143), pp. 407–412. doi: 10.1038/nature05915.

Bernstein, B. E., Kamal, M., Lindblad-toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., Mcmahon, S., Karlsson, E. K., Iii, E. J. K., Gingeras, T. R., Schreiber, S. L. and Lander, E. S. (2005) 'Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse', *Cell*, 120(2), pp. 169–181. doi: 10.1016/j.cell.2005.01.001.

Betts Lindroos, H., Ström, L., Itoh, T., Katou, Y., Shirahige, K. and Sjögren, C. (2006) 'Chromosomal Association of the Smc5/6 Complex Reveals that It Functions in Differently Regulated Pathways', *Molecular Cell*, 22(6), pp. 755–767. doi: 10.1016/j.molcel.2006.05.014.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J. and Hess, H. F. (2006) 'Imaging intracellular fluorescent proteins at nanometer resolution', *Science*, 313(5793), pp. 1642–1645. doi: 10.1126/science.1127344.

Bintu, B., Mateo, L. J., Su, J., Sinnott-armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N. and Zhuang, X. (2018) 'Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells', *Science*, 362. doi: 10.1126/science.aau1783.

Bird, A. (2002) ‘DNA methylation patterns and epigentic eemory’, *Genes and Development*, 16, pp. 16–21. doi: 10.1101/gad.947102.6.

Bird, A. P. (1987) ‘CpG islands as gene markers in the vertebrate nucleus’, *Trends in Genetics*, 3, pp. 342–347. doi: 10.1016/0168-9525(87)90294-0.

Blackwood, E. M. and Kadonaga, J. T. (1998) ‘Going the distance: A current view of enhancer action’, *Science*, 281(5373), pp. 60–63. doi: 10.1126/science.281.5373.60.

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C. T. and Zhuang, X. (2016) ‘Super-resolution imaging reveals distinct chromatin folding for different epigenetic states’, *Nature*, 529(7586), pp. 418–422. doi: 10.1038/nature16496.

Bonev, B., Cohen, N. M., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J., Tanay, A. and Cavalli, G. (2017) ‘Multiscale 3D Genome Rewiring during Mouse Neural Development’, *Cell*, 171(3), pp. 557–572. doi: 10.1016/j.cell.2017.09.043.

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, M. J. and Sherlock, G. (2004) ‘GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes’, *Bioinformatics*, 20(18), pp. 3710–3715. doi: 10.1093/bioinformatics/bth456.GO.

Branco, M. R. and Pombo, A. (2006) ‘Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations’, *PLoS Biology*, 4(5), pp. 780–788. doi: 10.1371/journal.pbio.0040138.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) ‘Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position’, *Nature Methods*, 10(12), pp. 1213–1218. doi: 10.1038/nmeth.2688.

Bulger, M. and Groudine, M. (1999) ‘Looping versus linking: Toward a model for long-distance gene activation’, *Genes and Development*, 13(19), pp. 2465–2477. doi: 10.1101/gad.13.19.2465.

Bulger, M. and Groudine, M. (2011) ‘Functional and mechanistic diversity of distal transcription enhancers’, *Cell*, 144(3), pp. 327–339. doi: 10.1016/j.cell.2011.01.024.

Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N. and Peters, J.-M. (2017) ‘Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl’, *Nature*. doi: 10.1038/nature22063.

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Dimond, A., Plagnol, V., Zerbino, D., Schoenfelder, S., Javierre, B.-M., Osborne, C., Fraser, P. and Spivakov, M. (2015) ‘CHiCAGO: Robust Detection of DNA Looping Interactions in Capture Hi-C data’, *Genome biology*, p. 028068. doi: 10.1101/028068.

Cairns, J., Orchard, W. R., Malysheva, V. and Spivakov, M. (2019) ‘Chicdiff: a computational pipeline for detecting differential chromosomal interactions in Capture Hi-C data’, *biorXiv*, 526269. doi: 10.1093/bioinformatics/btz450.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. feng and Fraser, P. (2002) ‘Long-range chromatin regulatory interactions in vivo’, *Nature Genetics*, 32(4), pp. 623–

626. doi: 10.1038/ng1051.

Cattoni, D. I., Gizzi, A. M. C., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbron, C., Déjardin, S., Fiche, J. B., González, I., Chang, J. M., Sexton, T., Marti-Renom, M. A., Bantignies, F., Cavalli, G. and Nollmann, M. (2017) 'Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions', *Nature Communications*, 8(1). doi: 10.1038/s41467-017-01962-x.

Chambeyron, S. and Bickmore, W. A. (2004) 'Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription', *Genes & Development*, 18(10), pp. 1119–1130. doi: 10.1101/gad.292104.tional.

Chathoth, K. T. and Zabet, N. R. (2019) 'Chromatin architecture reorganization during neuronal cell differentiation in Drosophila genome', *Genome Research*, 29(4), pp. 613–625. doi: 10.1101/gr.246710.118.

Chaumeil, J., Baccon, P. Le, Wutz, A. and Heard, E. (2006) 'A novel role for Xist RNA in the formation of a repressive nuclear', *Genes & development*, 20, pp. 2223–2237. doi: 10.1101/gad.380906.the.

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J. and Meyer, B. J. (2015) 'Condensin-driven remodelling of X chromosome topology during dosage compensation', *Nature*, 523(7559), pp. 240–244. doi: 10.1038/nature14450.

Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L. and Kirsch-Volders, M. (1982) 'Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments', *Human Genetics*, 62(3), pp. 201–209. doi: 10.1007/BF00333519.

Cremer, T. and Cremer, C. (2001) 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nature Reviews Genetics*, 2(April), pp. 292–301.

Cremer, T. and Cremer, M. (2010) 'Chromosome territories', *Cold Spring Harbor perspectives in biology*, 2.3, p. 38. doi: a003889.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. and Jaenisch, R. (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp. 21931–21936. doi: 10.1073/pnas.1016071107.

Cuartero, S., Weiss, F. D., Dharmalingam, G., Guo, Y., Ing-Simmons, E., Masella, S., Robles-Rebollo, I., Xiao, X., Wang, Y.-F., Barozzi, I., Djeghloul, D., Amano, M. T., Niskanen, H., Petretto, E., Dowell, R. D., Tachibana, K., Kaikkonen, M. U., Nasmyth, K. A., Lenhard, B., *et al.* (2018) 'Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation', *Nature Immunology*, 19(9), p. 932. doi: 10.1038/s41590-018-0184-1.

Dao, L. T. M. and Spicuglia, S. (2018) 'Transcriptional regulation by promoters with enhancer function', *Transcription*, 9(5), pp. 307–314. doi: 10.1080/21541264.2018.1486150.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) ‘Capturing chromosome conformation’, *Science (New York, N.Y.)*, 295(5558), pp. 1306–11. doi: 10.1126/science.1067799.

Dekker, J., Marti-Renom, M. a and Mirny, L. a (2013) ‘Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data’, *Nature reviews. Genetics*, 14(6), pp. 390–403. doi: 10.1038/nrg3454.

Dekker, J. and Mirny, L. (2016) ‘The 3D Genome as Moderator of Chromosomal Communication’, *Cell*, 164(6), pp. 1110–1121. doi: 10.1016/j.cell.2016.02.007.

Denslow, S. A. and Wade, P. A. (2007) ‘The human Mi-2/NuRD complex and gene regulation’, *Oncogene*, 26(37), pp. 5433–5438. doi: 10.1038/sj.onc.1210611.

Devaiah, B. N., Lewis, B. A., Cherman, N., Hewitt, M. C., Albrecht, B. K., Robey, P. G., Ozato, K., Sims, R. J. and Singer, D. S. (2012) ‘BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(18), pp. 6927–6932. doi: 10.1073/pnas.1120422109.

Devaiah, B. N., Case-Borden, C., Gegonne, A., Hsu, C. H., Chen, Q., Meerzaman, D., Dey, A., Ozato, K. and Singer, D. S. (2016) ‘BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin’, *Nature Structural and Molecular Biology*, 23(6), pp. 540–548. doi: 10.1038/nsmb.3228.

Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. and Ozato, K. (2003) ‘The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis’, *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), pp. 8758–8763. doi: 10.1073/pnas.1433065100.

Dileep, V., Ay, F., Sima, J., Vera, D. L., Noble, W. S. and Gilbert, D. M. (2015) ‘Topologically-associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program’, *Genome Research*, 25, pp. 1104–1113. doi: 10.1101/gr.183699.114.4.

Dillon, N. and Sabbattini, P. (2000) ‘Functional gene expression domains: Defining the functional unit of eukaryotic gene regulation’, *BioEssays*, 22(7), pp. 657–665. doi: 10.1002/1521-1878(200007)22:7<657::AID-BIES8>3.0.CO;2-2.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012) ‘Topological domains in mammalian genomes identified by analysis of chromatin interactions’, *Nature*, 485(7398), pp. 376–380. doi: 10.1038/nature11082.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A. and Ren, B. (2015) ‘Chromatin architecture reorganization during stem cell differentiation’, *Nature*, 518(7539), pp. 331–336. doi: 10.1038/nature14222.

Dixon, J. R., Gorkin, D. U. and Ren, B. (2016) ‘Chromatin Domains: The Unit of Chromosome Organization’, *Molecular Cell*, 62(5), pp. 668–680. doi: 10.1016/j.molcel.2016.05.018.

Donohoe, M. E., Zhang, L., Xu, N., Shi, Y. and Lee, J. T. (2007) ‘Identification of a

Ctcf Cofactor, Yy1, for the X Chromosome Binary Switch', *Molecular Cell*, 25, pp. 43–56. doi: 10.1016/j.molcel.2006.11.017.

Dorsett, D., Eissenberg, J. C., Misulovin, Z., Martens, A., Redding, B. and McKim, K. (2005) 'Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila', *Development*, 132(21), pp. 4743–4753. doi: 10.1242/dev.02064.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D. and Dekker, J. (2006) 'Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements', *Genome Research*, 16(10), pp. 1299–1309. doi: 10.1101/gr.5571506.

Dowen, J. M., Bilodeau, S., Orlando, D. A., Hübner, M. R., Abraham, B. J., Spector, D. L. and Young, R. A. (2013) 'Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements', *Stem Cell Reports*, 1(5), pp. 371–378. doi: 10.1016/j.stemcr.2013.09.002.

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S. and Aiden, E. L. (2016) 'Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom', *Cell Systems*, 3(1), pp. 99–101. doi: 10.1016/j.cels.2015.07.012.

Ebright, R. H. (2000) 'RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II', *Journal of Molecular Biology*, 304(5), pp. 687–698. doi: 10.1006/jmbi.2000.4309.

Edgar, R. (2002) 'Gene Expression Omnibus: NCBI gene expression and hybridization array data repository', *Nucleic Acids Research*, 30(1), pp. 207–210. doi: 10.1093/nar/30.1.207.

El-Sharnouby, S., Fischer, B., Magbanua, J. P., Umans, B., Flower, R., Choo, S. W., Russell, S. and White, R. (2017) 'Regions of very low H3K27me3 partition the Drosophila genome into topological domains', *PLoS ONE*, 12(3), pp. 1–23. doi: 10.1371/journal.pone.0172725.

Elizabeth Ing-Simmons, Seitan, V. C., Faure, A. J., Flicek, P., Carroll, T., Dekker, J., Fisher, A. G., Lenhard, B. and Merkenschlager, M. (2015) 'Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin', *Genome Research*, 25, pp. 504–513. doi: 10.1101/gr.184986.114.

Encode Consortium (2013) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. doi: 10.1038/nature11247.An.

Ewels, P., Krueger, F., Käller, M. and Andrews, S. (2017) 'Cluster Flow: A user-friendly bioinformatics workflow tool', *F1000Research*, 5(May), pp. 1–12. doi: 10.12688/f1000research.10335.1.

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B. R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I. and Mirny, L. (2018) 'Heterochromatin drives organization of conventional and inverted nuclei', *biorXiv*, pp. 1–19. doi: 10.1101/244038.

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G.,

Neiman, P. E., Collins, S. J. and Lobanenkov, V. V (1996) ‘An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes.’, *Molecular and Cellular Biology*, 16(6), pp. 2802–2813. doi: 10.1128/mcb.16.6.2802.

Flavahan, W. A., Drier, Y., Liau, B. B., Gillespie, S. M., Venteicher, A. S., Stemmer-Rachamimov, A. O., Suvà, M. L. and Bernstein, B. E. (2016) ‘Insulator dysfunction and oncogene activation in IDH mutant gliomas’, *Nature*, 33(8), pp. 839–841. doi: 10.1038/nbt.3301.Mammalian.

Floyd, S. R., Pacold, M. E., Huang, Q., Clarke, S. M., Lam, F. C., Cannell, I. G., Bryson, B. D., Rameseder, J., Lee, M. J., Blake, E. J., Fydrych, A., Ho, R., Greenberger, B. A., Chen, G. C., Maffa, A., Del Rosario, A. M., Root, D. E., Carpenter, A. E., Hahn, W. C., *et al.* (2013) ‘The bromodomain protein Brd4 insulates chromatin from DNA damage signalling’, *Nature*, 498(7453), pp. 246–250. doi: 10.1038/nature12147.

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V., Abdennur, N., Razin, S. V., Mirny, L. A. and Tachibana-Konwalski, K. (2017) ‘Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition’, *Nature*, 544(7648), pp. 110–114. doi: 10.1038/nature21711.

Fox, J. (2003) ‘Effect Displays for Generalized Linear Models’, *Journal of Statistical Software*, 8(15), pp. 1–27. doi: 10.2307/271037.

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., Spielmann, M., Timmermann, B., Wittler, L., Kurth, I., Cambiaso, P., Zuffardi, O., Houge, G., Lambie, L., Brancati, F., *et al.* (2016) ‘Formation of new chromatin domains determines pathogenicity of genomic duplications’, *Nature*, 538(7624), pp. 265–269. doi: 10.1038/nature19800.

Fraser, J., Ferrai, C., Chiariello, A. M., Schueler, M., Rito, T., Laudanno, G., Barbieri, M., Moore, B. L., Kraemer, D. C., Aitken, S., Xie, S. Q., Morris, K. J., Itoh, M., Kawaji, H., Jaeger, I., Hayashizaki, Y., Carninci, P., Forrest, A. R., Dostie, J., *et al.* (2015) ‘Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation’, *Mol Syst Biol*, 11, pp. 1–14. doi: 10.15252/msb.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) ‘Regularization Paths for Generalized Linear Regression, Ridge and Lasso’, *Journal of statistical software*, 33(1), p. 1.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L. A. (2016) ‘Formation of Chromosomal Domains by Loop Extrusion’, *Cell Reports*, 15(9), pp. 2038–2049. doi: 10.1016/j.celrep.2016.04.085.

Fujimoto, M., Takaki, E., Takii, R., Tan, K., Prakasam, R., Hayashida, N., Iemura, S. ichiro, Natsume, T. and Nakai, A. (2012) ‘RPA assists HSF1 access to nucleosomal DNA by recruiting histone chaperone FACT’, *Molecular Cell*, 48(2), pp. 182–194. doi: 10.1016/j.molcel.2012.07.026.

Fukaya, T., Lim, B. and Levine, M. (2016) ‘Enhancer Control of Transcriptional Bursting’, *Cell*, 166(2), pp. 358–368. doi: 10.1016/j.cell.2016.05.025.

Fulco, C. P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S. R., Perez, E.

M., Kane, M., Cleary, B., Lander, E. S. and Engreitz, J. M. (2016) ‘Systematic mapping of functional enhancer–promoter connections with CRISPR interference’, *Science*, 354(6313), pp. 769–773. doi: 10.1126/science.aag2445.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. Bin, Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G. Y., Huang, P. Y. H., Welboren, W. J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., *et al.* (2009) ‘An oestrogen-receptor-α-bound human chromatin interactome’, *Nature*, 462(7269), pp. 58–64. doi: 10.1038/nature08497.

Ganji, A. M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H. and Dekker, C. (2018) ‘Real-time imaging of DNA loop extrusion by condensin’, *Science*, 360(6384), pp. 102–105. doi: 10.1126/science.aar7831.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S. and Snyder, M. (2007) ‘What is a gene, post-ENCODE? History and updated definition’, *Genome research*, 17(6), pp. 669–681. doi: 10.1101/gr.6339607.Freely.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S. and Qi, L. S. (2013) ‘XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes’, *Cell*, 154(2), p. 442. doi: 10.1016/j.cell.2013.06.044.

Girdwood, D., Bumpass, D., Vaughan, O. A., Thain, A., Anderson, L. A., Snowden, A. W., Garcia-Wilson, E., Perkins, N. D. and Hay, R. T. (2003) ‘p300 Transcriptional Repression Is Mediated by SUMO Modification’, *Molecular Cell*, 11, pp. 1043–1054.

Glynn, E. F., Megee, P. C., Yu, H. G., Mistrot, C., Unal, E., Koshland, D. E., DeRisi, J. L. and Gerton, J. L. (2004) ‘Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae’, *PLoS Biology*, 2(9). doi: 10.1371/journal.pbio.0020259.

Goloborodko, A., Imakaev, M. V, Marko, J. F. and Mirny, L. (2016) ‘Compaction and segregation of sister chromatids via active loop extrusion’, *bioRxiv*, pp. 1–20. doi: 10.1101/038281.

Goodman, R. H. and Smolik, S. (2000) ‘CBP/p300 in cell growth, transformation, and development’, *Genes and Development*, 14(13), pp. 1553–1577.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. and Fraser, P. (2000) ‘Intergenic transcription and developmental remodeling of chromatin subdomains in the human β-globin locus’, *Molecular Cell*, 5(2), pp. 377–386. doi: 10.1016/S1097-2765(00)80432-3.

Gruber, S., Haering, C. H. and Nasmyth, K. (2003) ‘Chromosomal cohesin forms a ring’, *Cell*, 112(6), pp. 765–777. doi: 10.1016/S0092-8674(03)00162-4.

Grünberg, S. and Hahn, S. (2013) ‘Structural insights into transcription initiation by RNA polymerase II’, *Trends in Biochemical Sciences*, 38(12), pp. 603–611. doi: 10.1016/j.tibs.2013.09.002.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007) ‘A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells’,

Cell, 130(1), pp. 77–88. doi: 10.1016/j.cell.2007.05.042.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., Ren, B., Krainer, A. R., Maniatis, T. and Wu, Q. (2015) 'CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function', *Cell*, 162(4), pp. 900–910. doi: 10.1016/j.cell.2015.07.038.

Haarhuis, J. H. I., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P. H. L., Teunissen, H., Medema, R. H., van Steensel, B., Brummelkamp, T. R., de Wit, E. and Rowland, B. D. (2017) 'The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension', *Cell*, 169(4), pp. 693-707.e14. doi: 10.1016/j.cell.2017.04.013.

Haberle, V. and Lenhard, B. (2016) 'Promoter architectures and developmental gene regulation', *Seminars in Cell and Developmental Biology*, 57, pp. 11–23. doi: 10.1016/j.semcdb.2016.01.014.

Haberle, V. and Stark, A. (2018) 'Eukaryotic core promoters and the functional basis of transcription initiation', *Nature Reviews Molecular Cell Biology*, 19(10), pp. 621–637. doi: 10.1038/s41580-018-0028-8.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G. and Merkenschlager, M. (2009) 'Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus', *Nature*, 460(7253), pp. 410–413. doi: 10.1038/nature08079.

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., Gosden, M., Sharpe, J. A., Sloane-stanley, J. A., Hughes, J. R., Davies, B. and Higgs, D. R. (2017) 'Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo', *Nature cell biology*, 19(8), p. 952. doi: 10.1038/ncb3573.

Hargreaves, D. C., Horng, T. and Medzhitov, R. (2009) 'Control of Inducible Gene Expression by Signal-Dependent Transcriptional Elongation', *Cell*, 138(1), pp. 129–145. doi: 10.1016/j.cell.2009.05.047.

Harrison, P. (2015) 'Anscombe's 1948 variance stabilizing transformation for the negative binomial distribution is well suited to RNA-Seq expression data', in *F1000Research*.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Calcar, S. Van, Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. and Ren, B. (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics*, 39(3), pp. 311–318. doi: 10.1038/ng1966.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., *et al.* (2009) 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*, 459(7243), pp. 108–112. doi: 10.1038/nature07829.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X.,

Murre, C., Singh, H. and Glass, C. K. (2010) ‘Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities’, *Molecular Cell*, 38(4), pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Herzog, V. A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T. R., Wlotzka, W., Von Haeseler, A., Zuber, J. and Ameres, S. L. (2017) ‘Thiol-linked alkylation of RNA to assess expression dynamics’, *Nature Methods*, 14(12), pp. 1198–1204. doi: 10.1038/nmeth.4435.

Hirota, T., Gerlich, D., Koch, B., Ellenberg, J. and Peters, J. M. (2004) ‘Distinct functions of condensin I and II in mitotic chromosome assembly’, *Journal of Cell Science*, 117(26), pp. 6435–6445. doi: 10.1242/jcs.01604.

Ho Man Chan and Nicholas B. La Thangue (2001) ‘p300/CBP proteins: HATs for transcriptional bridges and scaffolds’, *Journal of cell science*, 114(13), pp. 2363–2373. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1757496.

Hsiung, C. C. S., Morrissey, C. S., Udugama, M., Frank, C. L., Keller, C. A., Baek, S., Giardine, B., Crawford, G. E., Sung, M. H., Hardison, R. C. and Blobel, G. A. (2015) ‘Genome accessibility is widely preserved and locally modulated during mitosis’, *Genome Research*, 25(2), pp. 213–225. doi: 10.1101/gr.180646.114.

Hsu, S. C., Gilgenast, T. G., Bartman, C. R., Edwards, C. R., Stonestrom, A. J., Huang, P., Emerson, D. J., Evans, P., Werner, M. T., Keller, C. A., Giardine, B., Hardison, R. C., Raj, A., Phillips-Cremins, J. E. and Blobel, G. A. (2017) ‘The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries’, *Molecular Cell*, 66(1), pp. 102-116.e7. doi: 10.1016/j.molcel.2017.02.027.

Huang, B., Yang, X.-D., Zhou, M.-M., Ozato, K. and Chen, L.-F. (2009) ‘Brd4 Coactivates Transcriptional Activation of NF-kB via Specific Binding to Acetylated RelA’, *Molecular and Cellular Biology*, 29(5), pp. 1375–1387. doi: 10.1128/mcb.01365-08.

Hughes, J. R., Roberts, N., Mcgowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R. and Higgs, D. R. (2014) ‘Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment’, *Nature Genetics*, 46(2), pp. 205–212. doi: 10.1038/ng.2871.

Huisinga, K. L., Brower-Toland, B. and Elgin, S. C. R. (2006) ‘The contradictory definitions of heterochromatin: Transcription and silencing’, *Chromosoma*, 115(2), pp. 110–122. doi: 10.1007/s00412-006-0052-x.

Ignatiadis, N., Klaus, B., Zaugg, J. B. and Huber, W. (2016) ‘Data-driven hypothesis weighting increases detection power in genome-scale multiple testing’, *Nature Methods*, 13(7), pp. 577–580. doi: 10.1038/nmeth.3885.

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J. and Mirny, L. A. (2012) ‘Iterative correction of Hi-C data reveals hallmarks of chromosome organization’, *Nature Methods*, 9(10), pp. 999–1003. doi: 10.1038/nmeth.2148.

International Human Genome Sequencing Consortium (2004) ‘Finishing the euchromatic sequence of the human genome’, *Nature*, 431(7011), pp. 931–945. doi:

10.1038/nature03001.

Jäger, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H. E., Heindl, A., Whiffin, N., Carnicer, M. J., Broome, L., Dryden, N., Nagano, T., Schoenfelder, S., Enge, M., Yuan, Y., Taipale, J., Fraser, P., Fletcher, O. and Houlston, R. S. (2015) 'Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci', *Nature Communications*, 6, p. 6178. doi: 10.1038/ncomms7178.

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., *et al.* (2016) 'Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters', *Cell*, 167(5), pp. 1369-1384.e19. doi: 10.1016/j.cell.2016.09.037.

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C. A., Schmitt, A. D., Espinoza, C. A. and Ren, B. (2013) 'A high-resolution map of the three-dimensional chromatin interactome in human cells', *Nature*, 503(7475), pp. 290–294. doi: 10.1038/nature12644.

Kadonaga, J. T. (2012) 'Perspectives on the RNA polymerase II core promoter', *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(1), pp. 40–51. doi: 10.1002/wdev.21.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J. and Young, R. A. (2010) 'Mediator and cohesin connect gene expression and chromatin architecture.', *Nature*, 467(7314), pp. 430–5. doi: 10.1038/nature09380.

Kalitsis, P., Zhang, T., Marshall, K. M., Nielsen, C. F. and Hudson, D. F. (2017) 'Condensin, master organizer of the genome', *Chromosome Research*, 25(1), pp. 61–76. doi: 10.1007/s10577-017-9553-0.

Kellum, R. and Schedl, P. (1991) 'A position-effect assay for boundaries of higher order chromosomal domains', *Cell*, 64(5), pp. 941–950. doi: 10.1016/0092-8674(91)90318-S.

Kellum, R. and Schedl, P. (1992) 'A Group of scs Elements Function as Domain Boundaries in an Enhancer-Blocking Assay', *Molecular and Cellular Biology*, 21(5), pp. 2424–2431.

El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.-R., Park, S., Wang, X., Krebs, J., Tripathi, S., Sakabe, N., Sobreira, D. R., Huang, S.-C., Rao, S. S. P., Pruett, N., Chauss, D., Sadler, E., *et al.* (2019) 'A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers', *Cell*, pp. 1–14. doi: 10.1016/j.cell.2019.07.011.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., De Vries, S. S., Janssen, H., Amendola, M., Nolen, L. D., Bickmore, W. A. and Van Steensel, B. (2013) 'Single-cell dynamics of genome-nuclear lamina interactions', *Cell*, 153(1), pp. 178–192. doi: 10.1016/j.cell.2013.02.028.

Landry, J. J. M., Pyl, P. T., Rausch, T., Zichner, T., Tekkedil, M. M., Stütz, A. M., Jauch, A., Aiyar, R. S., Pau, G., Delhomme, N., Gagneur, J., Korbel, J. O., Huber,

W. and Steinmetz, L. M. (2013) ‘The Genomic and Transcriptomic Landscape of a HeLa Cell Line’, *G3: Genes, Genomes, Genetics*, 3(8), pp. 1213–1224. doi: 10.1534/g3.113.005777.

Langmead, B. and Salzberg, S. L. (2012) ‘Fast gapped-read alignment with Bowtie 2’, *Nature Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2016) ‘Exact post-selection inference, with application to the lasso’, *Annals of Statistics*, 44(3), pp. 907–927. doi: 10.1214/15-AOS1371.

Lelij, P., Stocsits, R. R., Ladurner, R., Petzold, G., Kreidl, E., Koch, B., Schmitz, J., Neumann, B., Ellenberg, J. and Peters, J. (2014) ‘SNW1 enables sister chromatid cohesion by mediating the splicing of sororin and APC2 pre-mRNAs’, *The EMBO Journal*, 33(22), pp. 2643–2658. doi: 10.15252/embj.201488202.

Lengronne, A., Katou, Y., Mori, S., Yokabayashi, S., Kelly, G. P., Ito, T., Watanabe, Y., Shirahige, K. and Uhlmann, F. (2004) ‘Cohesin relocation from sites of chromosomal loading to places of convergent transcription’, *Nature*, 430(6999), pp. 573–578. doi: 10.1038/nature02742.

Lenhard, B., Sandelin, A. and Carninci, P. (2012) ‘Metazoan promoters: Emerging characteristics and insights into transcriptional regulation’, *Nature Reviews Genetics*, 13(4), pp. 233–245. doi: 10.1038/nrg3163.

Lettice, L. A., Horikoshi, T., Heaney, S. J. H., Van Baren, M. J., Van Der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., *et al.* (2002) ‘Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly’, *Proceedings of the National Academy of Sciences of the United States of America*, 99(11), pp. 7548–7553. doi: 10.1073/pnas.112212199.

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and de Graaff, E. (2003) ‘A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly’, *Human Molecular Genetics*, 12(14), pp. 1725–1735. doi: 10.1093/hmg/ddg180.

Li, B., Carey, M. and Workman, J. L. (2007) ‘The Role of Chromatin during Transcription’, *Cell*, 128(4), pp. 707–719. doi: 10.1016/j.cell.2007.01.015.

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J., Sim, H. S., Peh, S. Q., Mulawadi, F. H., Ong, C. T., Orlov, Y. L., Hong, S., Zhang, Z., Landt, S., Raha, D., *et al.* (2012) ‘Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation’, *Cell*, 148(1–2), pp. 84–98. doi: 10.1016/j.cell.2011.12.014.

Li, W., Hu, Y., Oh, S., Ma, Q., Merkurjev, D., Song, X., Zhou, X., Liu, Z., Tanasa, B., He, X., Chen, A. Y., Ohgi, K., Zhang, J., Liu, W. and Rosenfeld, M. G. (2015) ‘Condensin I and II Complexes License Full Estrogen Receptor α-Dependent Enhancer Activation’, *Molecular Cell*, 59(2), pp. 188–202. doi: 10.1016/j.molcel.2015.06.002.

Li, W., Notani, D. and Rosenfeld, M. G. (2016) ‘Enhancers as non-coding RNA

transcription units: recent insights and future perspectives', *Nature Reviews Genetics*, 17(4), pp. 207–223. doi: 10.1038/nrg.2016.4.

Lieberman-Aiden, E., Berkum, N. L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J. and Mirny, L. A. (2009) 'Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome', *Analyzer*, 33292(October), pp. 289–293. doi: 10.1038/nature08398.

Lill, N. L., Grossman, S. R., Ginsberg, D., DeCaprio, J. and Livingston, D. M. (1997) 'Binding and modulation of p53 by p300/CBP coactivators', *Nature*, 387(6635), pp. 823–827. doi: 10.1038/42981.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N. B., Vega, H., Jackson, L. G., Shirahige, K. and Krantz, I. D. (2009) 'Transcriptional Dysregulation in NIPBL and Cohesin Mutant Human Cells', 7(5). doi: 10.1371/journal.pbio.1000119.

Lomvardas, S. and Thanos, D. (2002) 'Modifying gene expression programs by altering core promoter chromatin architecture', *Cell*, 110(2), pp. 261–271. doi: 10.1016/S0092-8674(02)00822-X.

Long, H. K., Prescott, S. L. and Wysocka, J. (2016) 'Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution', *Cell*, 167(5), pp. 1170–1187. doi: 10.1016/j.cell.2016.09.018.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., *et al.* (2015) 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell*, 161(5), pp. 1012–1025. doi: 10.1016/j.cell.2015.04.004.

Maass, P. G., Barutcu, A. R., Weiner, C. L. and Rinn, J. L. (2018) 'Inter-chromosomal Contact Properties in Live-Cell Imaging and in Hi-C', *Molecular Cell*, 69(6), pp. 1039-1045.e3. doi: 10.1016/j.molcel.2018.02.007.

Malik, S. and Roeder, R. G. (2010) 'The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation', *Nature Reviews Genetics*, 11(11), pp. 761–772. doi: 10.1038/nrg2901.

Margueron, R. and Reinberg, D. (2011) 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469(7330), pp. 343–349. doi: 10.1038/nature09784.

Maston, G. A., Evans, S. K. and Green, M. R. (2006) 'Transcriptional Regulatory Elements in the Human Genome', *Annual Review of Genomics and Human Genetics*, 7(1), pp. 29–59. doi: 10.1146/annurev.genom.7.080505.115623.

Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y. and Taatjes, D. J. (2010) 'P53 activates transcription by directing structural shifts in Mediator', *Nature Structural*

and Molecular Biology, 17(6), pp. 753–760. doi: 10.1038/nsmb.1816.

Michaud, J., Praz, V., Faresse, N. J., JnBaptiste, C. K., Tyagi, S., Schütz, F. and Herr, W. (2013) 'HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy', *Genome Research*, 23(6), pp. 907–916. doi: 10.1101/gr.150078.112.

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G. A., Fraser, P., Luscombe, N. M. and Osborne, C. S. (2015) 'Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C', *Nature Genetics*, 47(6), pp. 598–606. doi: 10.1038/ng.3286.

Millau, J. F. and Gaudreau, L. (2011) 'CTCF, cohesin, and histone variants: Connecting the genome', *Biochemistry and Cell Biology*, 89(5), pp. 505–513. doi: 10.1139/o11-052.

Moore, B. L., Aitken, S. and Semple, C. A. (2015) 'Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization', *Genome Biology*, 16(1), pp. 1–14. doi: 10.1186/s13059-015-0661-x.

Morawska, M. and Ulrich, H. D. (2013) 'An expanded tool kit for the auxin-inducible degron system in budding yeast', *Yeast*, 30(9), pp. 341–351. doi: 10.1002/yea.

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J. and Chang, H. Y. (2016) 'HiChIP: Efficient and sensitive analysis of protein-directed genome architecture', *Nature Methods*, 13(11), pp. 919–922. doi: 10.1038/nmeth.3999.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A. and Fraser, P. (2013) 'Single-cell Hi-C reveals cell-to-cell variability in chromosome structure', *Nature*, 502(7469), pp. 59–64. doi: 10.1038/nature12593.

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson Cohen, N., Wingett, S., Fraser, P. and Tanay, A. (2017) 'Cell-cycle dynamics of chromosomal organization at single-cell resolution', *Nature*, 547(7661), pp. 61–67. doi: 10.1038/nature23001.

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O. and Reinberg, D. (2015) 'CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation', *Science*, 347(6225), pp. 1017–1021. doi: 10.1126/science.1262088.

Narendra, V., Bulajié, M., Dekker, J., Mazzoni, E. O. and Reinberg, D. (2016) 'CTCF-mediated topological boundaries during development foster appropriate gene regulation', *Genes and Development*, 30(24), pp. 2657–2662. doi: 10.1101/gad.288324.116.

Nasmyth, K. and Haering, C. H. (2009) 'Cohesin: Its Roles and Mechanisms', *Annual Review of Genetics*, 43(1), pp. 525–558. doi: 10.1146/annurev-genet-102108-134233.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. a and Dekker, J. (2013) 'Organization of the mitotic chromosome', *Science*, 342(6161), pp. 948–53. doi: 10.1126/science.1236083.

Nir, G., Farabella, I., Estrada, C. P., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., Lee, S. H., Nguyen, S. C., McCole, R. B., Chattoraj, S., Erceg, J., Abed, J. A., M., N., Martins, C., Nguyen, H. Q., Hannan, M. A., Russell, S., Durand, N. C., Rao, S. S. P., *et al.* (2018) 'Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling.', *PLoS Genetics*, 14(12), p. e1007872.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A. A., Mechtler, K. and Peters, J. M. (2010) 'Sororin mediates sister chromatid cohesion by antagonizing Wapl', *Cell*, 143(5), pp. 737–749. doi: 10.1016/j.cell.2010.10.031.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. and Heard, E. (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*, 485, pp. 381–385. doi: 10.1038/nature11049.

Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A. and Bruneau, B. G. (2017) 'Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization', *Cell*, 169(5), pp. 930-944.e22. doi: 10.1016/j.cell.2017.05.004.

Norton, H. K., Emerson, D. J., Huang, H., Kim, J., Titus, K. R., Gu, S., Bassett, D. S. and Phillips-Cremins, J. E. (2018) 'Detecting hierarchical genome folding with network modularity', *Nature Methods*, 15(2), pp. 119–122. doi: 10.1038/nmeth.4560.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. and Mirny, L. A. (2018) 'Chromatin organization by an interplay of loop extrusion and compartmental segregation', *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), pp. E6697–E6706. doi: 10.1073/pnas.1717730115.

Ogbourne, S. and Antalis, T. M. (1998) 'Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes', *Biochemical Journal*, 331(1), pp. 1–14. doi: 10.1042/bj3310001.

Ong, C. T. and Corces, V. G. (2014) 'CTCF: An architectural protein bridging genome topology and function', *Nature Reviews Genetics*, 15(4), pp. 234–246. doi: 10.1038/nrg3663.

Ou, J., Liu, H., Yu, J., Kelliher, M. A., Castilla, L. H., Lawson, N. D. and Zhu, L. J. (2018) 'ATACseqQC: A Bioconductor package for post-alignment quality assessment of ATAC-seq data', *BMC Genomics*, 19(1), p. 169. doi: 10.1186/s12864-018-4559-3.

P.Moreau, R.Hen, B.Wasylyk, R.Everett, M.P.Gaub and P.Chambon (1981) 'The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants', *Nucleic Acids Research*, 9(22), pp. 6047–6068.

Palozola, K. C., Donahue, G., Liu, H., Grant, G. R., Becker, J. S., Cote, A., Yu, H., Raj, A. and Zaret, K. S. (2017) 'Mitotic transcription and waves of gene reactivation during mitotic exit', *Science*, 4671(September), pp. 1–9. Available at: http://science.sciencemag.org/content/early/2017/09/13/science.aal4671.full.

Papai, G., Tripathi, M. K., Ruhlmann, C., Layer, J. H., Weil, P. A. and Schultz, P. (2010) 'TFIIA and the transactivator Rap1 cooperate to commit TFIID for

transcription initiation', *Nature*, 465(7300), pp. 956–960. doi: 10.1038/nature09080.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A. G. and Merkenschlager, M. (2008) 'Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms', *Cell*, 132(3), pp. 422–433. doi: 10.1016/j.cell.2008.01.011.

Patel, M. C., Debrosse, M., Smith, M., Dey, A., Huynh, W., Sarai, N., Heightman, T. D., Tamura, T. and Ozato, K. (2013) 'BRD4 Coordinates Recruitment of Pause Release Factor P-TEFb and the Pausing Complex NELF/DSIF To Regulate Transcription Elongation of Interferon-Stimulated Genes', *Molecular and Cellular Biology*, 33(12), pp. 2497–2507. doi: 10.1128/mcb.01180-12.

Pauli, A., Bemmel, J. G. Van, Oliveira, R. A., Itoh, T., Shirahige, K., Steensel, B. Van and Nasmyth, K. (2010) 'A Direct Role for Cohesin in Gene Regulation and Ecdysone Response in Drosophila Salivary Glands', *Current Biology*, 20(20), pp. 1787–1798. doi: 10.1016/j.cub.2010.09.006.

Peters, J. M., Tedeschi, A. and Schmitz, J. (2008) 'The cohesin complex and its roles in chromosome biology', *Genes and Development*, 22(22), pp. 3089–3114. doi: 10.1101/gad.1724308.

Phatnani, H. P. and Greenleaf, A. L. (2006) 'Phosphorylation and functions of the RNA polymerase II CTD', *Genes and Development*, 20(21), pp. 2922–2936. doi: 10.1101/gad.1477006.

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gülsoy, G., Dennis, J. H., Snyder, M. P., Stamatoyannopoulos, J. A., Taylor, J., Hardison, R. C., Kahveci, T., *et al.* (2014) 'Topologically associating domains are stable units of replication-timing regulation', *Nature*, 515(7527), pp. 402–405. doi: 10.1038/nature13986.

Poss, Z. C., Ebmeier, C. C. and Taatjes, D. J. (2013) 'The Mediator complex and transcription regulation', *Critical Reviews in Biochemistry and Molecular Biology*, 48(6), pp. 575–608. doi: 10.3109/10409238.2013.840259.

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C. A., Schneider, R. and Bickmore, W. A. (2016) 'Histone H3 globular domain acetylation identifies a new class of enhancers', *Nature Genetics*, 48(6), pp. 681–686. doi: 10.1038/ng.3550.

Pünzeler, S., Link, S., Wagner, G., Keilhauer, E. C., Kronbeck, N., Spitzer, R. M., Leidescher, S., Markaki, Y., Mentele, E., Regnard, C., Schneider, K., Takahashi, D., Kusakabe, M., Vardabasso, C., Zink, L. M., Straub, T., Bernstein, E., Harata, M., Leonhardt, H., *et al.* (2017) 'Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation', *The EMBO Journal*, 36(15), pp. 2263–2279. doi: 10.15252/embj.201695757.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Racko, D., Benedetti, F., Goundaroulis, D. and Stasiak, A. (2018) 'Chromatin Loop Extrusion and Chromatin Unknotting', *Polymers*, 10(10), pp. 1126–1137. doi: 10.3390/polym10101126.

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., Syed, T., Emons, B. J. M., Gifford, D. K. and Sherwood, R. I. (2016) 'High-throughput mapping of regulatory DNA', *Nature Biotechnology*, 34(2), pp. 167–174. doi: 10.1038/nbt.3468.

Ramani, V., Deng, X., Qiu, R., Gunderson, K. L., Steemers, F. J., Disteche, C. M., Noble, W. S., Duan, Z. and Shendure, J. (2017) 'Massively multiplex single-cell Hi-C', *Nature Methods*, 14(3), pp. 263–266. doi: 10.1038/nmeth.4155.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013) 'Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity', *Cell*, 154(6), pp. 1380–1389. doi: 10.1016/j.cell.2013.08.021.

Rao, S. S. P., Huang, S.-C., Hilaire, B. G. S., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171(2), pp. 305-320.e24. doi: 10.1016/j.cell.2017.09.026.

Rao, S. S. P. S. P., Huntley, M. H. H., Durand, N. C. C., Stamenova, E. K. K., Bochkov, I. D. D., Robinson, J. T. T., Sanborn, A. L. L., Machol, I., Omer, A. D. D., Lander, E. S. S. and Aiden, E. L. L. (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*, 159(7), pp. 1665–1680. doi: 10.1016/j.cell.2014.11.021.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P. and Hendrich, B. (2012) 'NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression', *EMBO Journal*, 31(3), pp. 593–605. doi: 10.1038/emboj.2011.431.

Rhodes, J. D. P., Feldmann, A., Hernández-Rodríguez, B., Díaz, N., Brown, J. M., Fursova, N. A., Blackledge, N. P., Prathapan, P., Dobrinic, P., Huseyin, M., Szczurek, A., Kruse, K., Nasmyth, K. A., Buckle, V. J., Vaquerizas, J. M. and Klose, R. J. (2019) 'Cohesin disrupts polycomb-dependent chromosome interactions', *bioRxiv*, p. 593970. doi: 10.1101/593970.

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-Idrissi, I., Nguyen Huynh, T. H., Ozadam, H., Dekker, J. and Duboule, D. (2017) 'The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes', *Genes and Development*, 31(22), pp. 2264–2281. doi: 10.1101/gad.307769.117.

Roeder, R. G. and Rutter, W. J. (1969) 'Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms', *Nature*, 224, pp. 234–237.

Roeder, R. G. and Rutter, W. J. (1970) 'Specific nucleolar and nucleoplasmic RNA polymerases.', *Proceedings of the National Academy of Sciences of the United States of America*, 65(3), pp. 675–682. doi: 10.1073/pnas.65.3.675.

Rollins, R. A., Korom, M., Aulner, N., Martens, A. and Dorsett, D. (2004) 'Drosophila Nipped-B Protein Supports Sister Chromatid Cohesion and Opposes the Stromalin/Scc3 Cohesion Factor To Facilitate Long-Range Activation of the cut

Gene', *Molecular and Cellular Biology*, 24(8), pp. 3100–3111. doi: 10.1128/mcb.24.8.3100-3111.2004.

Rosmarin, A. G., Resendes, K. K., Yang, Z., McMillan, J. N. and Fleming, S. L. (2004) 'GA-binding protein transcription factor: A review of GABP as an integrator of intracellular signaling and protein-protein interactions', *Blood Cells, Molecules, and Diseases*, 32(1), pp. 143–154. doi: 10.1016/j.bcmd.2003.09.005.

Rust, M. J., Bates, M. and Zhuang, X. (2006) 'Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)', *Nature Methods*, 3(10), p. 793. doi: 10.1038/nmeth929.

Saksouk, N., Simboeck, E. and Déjardin, J. (2015) 'Constitutive heterochromatin formation and transcription in mammals', *Epigenetics and Chromatin*, 8(1), pp. 1–17. doi: 10.1186/1756-8935-8-3.

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S. and Aiden, E. L. (2015) 'Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes', *Proceedings of the National Academy of Sciences* . doi: 10.1073/pnas.1518552112.

Sankar, N., Baluchamy, S., Kadeppagari, R. K., Singhal, G., Weitzman, S. and Thimmapaya, B. (2008) 'p300 provides a corepressor function by cooperating with YY1 and HDAC3 to repress c-Myc', *Oncogene*, 27(43), pp. 5717–5728. doi: 10.1038/onc.2008.181.

Schaaf, C. A., Misulovin, Z., Sahota, G., Siddiqui, A. M., Schwartz, Y. B., Tatyana, G., Pirrotta, V., Gause, M. and Dorsett, D. (2009) 'Regulation of the Drosophila Enhancer of split and invected-engrailed Gene Complexes by Sister Chromatid Cohesion Proteins', *PloS one*, 4(7). doi: 10.1371/journal.pone.0006202.

Schardin, M., Cremer, T., Hager, H. D. and Lang, M. (1985) 'Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories', *Human Genetics*, 71(4), pp. 281–287. doi: 10.1007/BF00388452.

Schmitt, A. D., Hu, M. and Ren, B. (2016) 'Genome-wide mapping and analysis of chromosome architecture', *Nature Reviews Molecular Cell Biology*, 17(12), pp. 743–755. doi: 10.1038/nrm.2016.104.

Schoenfelder, S., Furlan-magaril, M., Mifsud, B., Tavares-cadete, F., Sugar, R., Javierre, B., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S. W., Dimitrova, E., Dimond, A., Edelman, L. B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., *et al.* (2015) 'The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements', *Genome Res.*, pp. 1–16. doi: 10.1101/gr.185272.114.Freely.

Schoenfelder, S., Javierre, B. M., Furlan-Magaril, M., Wingett, S. W. and Fraser, P. (2018) 'Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions', *J. Vis. Exp.*, 136(e57320).

Schoenfelder, S. and Fraser, P. (2019) 'Long-range enhancer–promoter contacts in gene expression control', *Nature Reviews Genetics*. doi: 10.1038/s41576-019-0128-

0.

Schwalie, P. C., Ward, M. C., Cain, C. E., Faure, A. J., Gilad, Y., Odom, D. T. and Flicek, P. (2013) ‘Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes’, *Genome Biology*, 14(12), p. R148. doi: 10.1186/gb-2013-14-12-r148.

Schwartz, S., Meshorer, E. and Ast, G. (2009) ‘Chromatin organization marks exon-intron structure’, *Nature Structural and Molecular Biology*, 16(9), pp. 990–995. doi: 10.1038/nsmb.1659.

Schwartz, Y. B. and Pirrotta, V. (2007) ‘Polycomb silencing mechanisms and the management of genomic programmes’, *Nature Reviews Genetics*, 8(1), pp. 9–22. doi: 10.1038/nrg1981.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C., Mirny, L. and Spitz, F. (2017) ‘Two independent modes of chromosome organization are revealed by cohesin removal’, *Nature*, 551(7678), p. 094185. doi: 10.1101/094185.

Seitan, V. C., Faure, A. J., Zhan, Y., McCord, R. P., Lajoie, B. R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A. G., Flicek, P., Dekker, J. and Merkenschlager, M. (2013) ‘Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments’, *Genome Research*, 23(12), pp. 2066–2077. doi: 10.1101/gr.161620.113.

Shen, Y., Yue, F., Mc Cleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Ren, B. and Lobanenkov, V. V. (2012) ‘A map of the cis-regulatory sequences in the mouse genome’, *Nature*, 488(7409), pp. 116–120. doi: 10.1038/nature11243.

Shlyueva, D., Stampfel, G. and Stark, A. (2014) ‘Transcriptional enhancers: From properties to genome-wide predictions’, *Nature Reviews Genetics*, 15(4), pp. 272–286. doi: 10.1038/nrg3682.

Shrivastava, A. and Calame, K. (1994) ‘An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1’, *Nucleic Acids Research*, 22(24), pp. 5151–5155. doi: 10.1093/nar/22.24.5151.

Siggens, L., Cordeddu, L., Rönnerblad, M., Lennartsson, A. and Ekwall, K. (2015) ‘Transcription-coupled recruitment of human CHD1 and CHD2 influences chromatin accessibility and histone H3 and H3.3 occupancy at active chromatin regions’, *Epigenetics and Chromatin*, 8(1), pp. 1–14. doi: 10.1186/1756-8935-8-4.

Sigova, A. A., Abraham, B. J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y. E., Jangi, M., Giallourakis, C. C., Sharp, P. A. and Young, R. A. (2015) ‘Transcription factor trapping by RNA in gene regulatory elements’, *Science*, 350(6263), pp. 978–982.

Sikorski, T. W., Ficarro, S. B., Holik, J., Kim, T. S., Rando, O. J., Marto, J. A. and Buratowski, S. (2011) ‘Sub1 and RPA associate with RNA polymerase II at different stages of transcription’, *Molecular Cell*, 44(3), pp. 397–409. doi: 10.1016/j.molcel.2011.09.013.

Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K. N., Holcomb, N. P., Turner, J. L., Paulsen, M. T., Rivera-Mulia, J. C., Trevilla-Garcia, C., Bartlett, D. A.,

Zhao, P. A., Washburn, B. K., Nora, E. P., Kraft, K., Mundlos, S., Bruneau, B. G., Ljungman, M., Fraser, P., *et al.* (2019) 'Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication', *Cell*, 176(4), pp. 816-830.e18. doi: 10.1016/j.cell.2018.11.036.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S. M., Schroth, G. P., Tanay, A. and Hadjur, S. (2013) 'Cohesin-mediated interactions organize chromosomal domain architecture', *The EMBO Journal*, 32(24), pp. 3119–3129. doi: 10.1038/emboj.2013.237.

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J. and Joffe, B. (2009) 'Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution', *Cell*, 137(2), pp. 356–368. doi: 10.1016/j.cell.2009.01.052.

Solovei, I., Wang, A. S., Thanisch, K., Schmidt, C. S., Krebs, S., Zwerger, M., Cohen, T. V., Devys, D., Foisner, R., Peichl, L., Herrmann, H., Blum, H., Engelkamp, D., Stewart, C. L., Leonhardt, H. and Joffe, B. (2013) 'LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation', *Cell*, 152(3), pp. 584–598. doi: 10.1016/j.cell.2013.01.009.

Song, W., Sharan, R. and Ovcharenko, I. (2019) 'The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance', *Genome Biology*, 20(1), p. 197.

Soutourina, J. (2018) 'Transcription regulation by the Mediator complex', *Nature Reviews Molecular Cell Biology*, 19(4), pp. 262–274. doi: 10.1038/nrm.2017.115.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors: From enhancer binding to developmental control', *Nature Reviews Genetics*, 13(9), pp. 613–626. doi: 10.1038/nrg3207.

Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P. and de Laat, W. (2012) 'Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation', *Methods*, 58(3), pp. 221–230. doi: 10.1016/j.ymeth.2012.04.009.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. and Spitz, F. (2014) 'Functional and topological characteristics of mammalian regulatory domains', *Genome Research*, 24(3), pp. 390–400. doi: 10.1101/gr.163519.113.

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M., Bantignies, F. and Cavalli, G. (2018) 'TADs are 3D structural units of higher-order chromosome organization in Drosophila', *Science Advances*, 4(2), pp. 1–14. doi: 10.1126/sciadv.aar8082.

Taatjes, D. J., Schneider-Poetsch, T. and Tjian, R. (2004) 'Distinct conformational states of nuclear receptor-bound CRSP-Med complexes', *Nature Structural and Molecular Biology*, 11(7), pp. 664–671. doi: 10.1038/nsmb789.

Tarjan, D. R., Flavahan, W. A. and Bernstein, B. E. (2019) 'Epigenome editing strategies for the functional annotation of CTCF insulators', *Nature Communications*, 10(1), pp. 1–8. doi: 10.1038/s41467-019-12166-w.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E.,

Davidson, I. F., Tang, W., Cisneros, D. a, Bhaskara, V., Nishiyama, T., Vaziri, A., Wutz, A., Ellenberg, J. and Peters, J.-M. (2013) 'Wapl is an essential regulator of chromatin structure and chromosome segregation.', *Nature*, 501(7468), pp. 564–8. doi: 10.1038/nature12471.

The ENCODE Project Consortium (2011) 'A user's guide to the Encyclopedia of DNA elements (ENCODE)', *PLoS Biology*, 9(4). doi: 10.1371/journal.pbio.1001046.

Thomas, M. C. and Chiang, C. M. (2006) 'The general transcription machinery and general cofactors.', *Critical reviews in biochemistry and molecular biology*, 41(3), pp. 105–178. doi: 10.1080/10409230600648736.

Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E. and Baylin, S. B. (2008) 'A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations', *Genome Research*, 18(7), pp. 1171–1179. doi: 10.1101/gr.073452.107.

Todeschini, A. L., Georges, A. and Veitia, R. A. (2014) 'Transcription factors: Specific DNA binding and specific gene regulation', *Trends in Genetics*, 30(6), pp. 211–219. doi: 10.1016/j.tig.2014.04.002.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and De Laat, W. (2002) 'Looping and interaction between hypersensitive sites in the active β-globin locus', *Molecular Cell*, 10(6), pp. 1453–1465. doi: 10.1016/S1097-2765(02)00781-5.

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C. and Spitz, F. (2014) 'Long-range enhancers regulating Myc expression are required for normal facial morphogenesis', *Nature Genetics*, 46(7), pp. 753–758. doi: 10.1038/ng.2971.

Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M. V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E. J., Reimand, J., Hadjur, S., Gingras, A.-C. and Wilson, M. D. (2016) 'Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders', *Genome Biol*, 17(1), pp. 1–22. doi: 10.1186/s13059-016-1043-8.

Valton, A. L. and Dekker, J. (2016) 'TAD disruption as oncogenic driver', *Current Opinion in Genetics and Development*, 36, pp. 34–40. doi: 10.1016/j.gde.2016.03.008.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. and Luscombe, N. M. (2009) 'A census of human transcription factors: Function, expression and evolution', *Nature Reviews Genetics*, 10(4), pp. 252–263. doi: 10.1038/nrg2538.

Vastenhouw, N. L. and Schier, A. F. (2012) 'Bivalent histone modifications in early embryogenesis', *Current Opinion in Cell Biology*, 24(3), pp. 374–386. doi: 10.1016/j.ceb.2012.03.009.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth. Available at: http://www.stats.ox.ac.uk/pub/MASS4.

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G. and Mann, M. (2010) 'Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers', *Cell*, 142(6), pp. 967–980. doi:

10.1016/j.cell.2010.08.020.

Versteeg, R., van Schaik, B. D. C., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H. J. and van Kampen, A. H. C. (2003) 'The human transcriptome map reveals extremes in gene dentistry, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes', *Genome Research*, 13(9), pp. 1998–2004. doi: 10.1101/gr.1649303.

Vian, L., Pękowska, A., Rao, S. S. P., Kieffer-Kwon, K. R., Jung, S., Baranello, L., Huang, S. C., El Khattabi, L., Dose, M., Pruett, N., Sanborn, A. L., Canela, A., Maman, Y., Oksanen, A., Resch, W., Li, X., Lee, B., Kovalchuk, A. L., Tang, Z., *et al.* (2018) 'The Energetics and Physiological Impact of Cohesin Extrusion', *Cell*, 173(5), pp. 1165-1178.e20. doi: 10.1016/j.cell.2018.03.072.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A. and Hadjur, S. (2015) 'Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture', *Cell Reports*, 10(8), pp. 1297–1309. doi: 10.1016/j.celrep.2015.02.004.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M. and Pennacchio, L. A. (2009) 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature*, 457(7231), pp. 854–858. doi: 10.1038/nature07730.

Visel, A., Rubin, E. M. and Pennacchio, L. A. (2009) 'Genomic views of distant-acting enhancers', *Nature*, 461(7261), pp. 199–205. doi: 10.1038/nature08451.

Vispé, S., Devries, L., Créancier, L., Besse, J., Bréand, S., Hobson, D. J., Svejstrup, J. Q., Annereau, J., Cussac, D., Dumontet, C., Guilbaud, N., Barret, J. and Bailly, C. (2009) 'Triptolide is an inhibitor of RNA polymerase I and II – dependent transcription leading predominantly to down-regulation of short-lived mRNA', *Molecular cancer therapeutics*, 8(10), pp. 2780–2790. doi: 10.1158/1535-7163.MCT-09-0549.

Wallace, J. A. and Felsenfeld, G. (2007) 'We gather together: insulators and genome organization', *Current Opinion in Genetics and Development*, 17(5), pp. 400–407. doi: 10.1016/j.gde.2007.08.005.

Wang, Q., Carroll, J. S. and Brown, M. (2005) 'Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking', *Molecular Cell*, 19(5), pp. 631–642. doi: 10.1016/j.molcel.2005.07.018.

Wang, R., Li, Q., Helfer, C. M., Jiao, J. and You, J. (2012) 'Bromodomain protein Brd4 associated with acetylated chromatin is important for maintenance of higher-order chromatin structure', *Journal of Biological Chemistry*, 287(14), pp. 10738–10752. doi: 10.1074/jbc.M111.323493.

Wang, Y., Lu, J., He, L. and Yu, Q. (2011) 'Triptolide (TPL) inhibits global transcription by inducing proteasome-dependent degradation of RNA polymerase II (Pol II)', *PloS one*, 6(9), p. e23993. doi: 10.1371/journal.pone.0023993.

Warner, J. R. (1999) 'The economics of ribosome biosynthesis in yeast', *Trends in Biochemical Sciences*, 24(11), pp. 437–440.

Weinmann, R. and Roeder, R. G. (1974) 'Role of DNA-Dependent RNA Polymerase

III in the Transcription of the tRNA and 5S RNA Genes', *Proceedings of the National Academy of Sciences of the United States of America*, 71(5), pp. 1790–1794. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC388326/pdf/pnas00058-0212.pdf.

Weinreb, C. and Raphael, B. J. (2016) 'Identification of hierarchical chromatin domains', *Bioinformatics*, 32(11), pp. 1601–1609. doi: 10.1093/bioinformatics/btv485.

Weintraub, A. S., Li, C. H., Zamudio, A. V, Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S. and Young, R. A. (2017) 'YY1 Is a Structural Regulator of Enhancer-Promoter Loops', *Cell*, 171(7), pp. 1573–1588. doi: 10.1016/j.cell.2017.11.008.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J.-M. (2008) 'Cohesin mediates transcriptional insulation by CCCTC-binding factor.', *Nature*, 451(7180), pp. 796–801. doi: 10.1038/nature06634.

Whalen, S., Truty, R. M. and Pollard, K. S. (2016) 'Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin', *Nature Genetics*, 48(5), p. 488. doi: 10.1038/ng.3539.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. and Young, R. A. (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153(2), pp. 307–319. doi: 10.1016/j.cell.2013.03.035.

Wijchers, P. J., Krijger, P. H. L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M. J. A. M., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., Anink-Groenen, L. C. M., Verschure, P. J. and de Laat, W. (2016) 'Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments', *Molecular Cell*, 61(3), pp. 461–473. doi: 10.1016/j.molcel.2016.01.001.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P. and Andrews, S. (2015) 'HiCUP: Pipeline for mapping and processing Hi-C data', *F1000Research*, 4, pp. 1–12. doi: 10.12688/f1000research.7334.1.

Wingett, S. W. and Andrews, S. (2018) 'Fastq screen: A tool for multi-genome mapping and quality control', *F1000Research*, 7(0), pp. 1–13. doi: 10.12688/f1000research.15931.1.

Wissink, E. M., Vihervaara, A., Tippens, N. D. and Lis, J. T. (2019) 'Nascent RNA analyses: tracking transcription and its regulation', *Nature Reviews Genetics*, 17, pp. 19–21. doi: 10.1038/s41576-019-0159-6.

Wit, E. De, Vos, E. S. M., Holwerda, S. J. B., Valdes-quezada, C., Verstegen, M. J. A. M., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H. L. and Laat, W. De (2015) 'CTCF Binding Polarity Determines Chromatin Looping', *Molecular Cell*, 60, pp. 1–9. doi: 10.1016/j.molcel.2015.09.023.

Wold, M. S. (1997) 'REPLICATION PROTEIN A:A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism', *Annual Review of Biochemistry*, 66(1), pp. 61–92. doi: 10.1146/annurev.biochem.66.1.61.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M. J., Walther, N., Koch, B., Kueblbeck, M., Ellenberg, J., Zuber, J., Fraser, P. and Peters, J. (2017) 'Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF , WAPL , and PDS 5 proteins', *The EMBO Journal*, 36(24), pp. 3573–3599. doi: 10.15252/embj.201798004.

Yan, J., Chen, S.-A. A., Local, A., Liu, T., Qiu, Y., Dorighi, K. M., Preissl, S., Rivera, C. M., Wang, C., Ye, Z., Ge, K., Hu, M., Wysocka, J. and Ren, B. (2018) 'Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers', *Cell Research*, 28(2), pp. 204–220. doi: 10.1038/cr.2018.1.

Yan, K. K., Lou, S. and Gerstein, M. (2017) 'MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions', *PLoS Computational Biology*, 13(7), pp. 1–22. doi: 10.1371/journal.pcbi.1005647.

Yang, Z. F., Mott, S. and Rosmarin, A. G. (2007) 'The Ets transcription factor GABP is required for cell-cycle progression', *Nature Cell Biology*, 9(3), pp. 339–346. doi: 10.1038/ncb1548.

Yu, H., Mashtalir, N., Daou, S., Hammond-Martel, I., Ross, J., Sui, G., Hart, G. W., Rauscher, F. J., Drobetsky, E., Milot, E., Shi, Y. and Affar, E. B. (2010) 'The Ubiquitin Carboxyl Hydrolase BAP1 Forms a Ternary Complex with YY1 and HCF-1 and Is a Critical Regulator of Gene Expression', *Molecular and Cellular Biology*, 30(21), pp. 5071–5085. doi: 10.1128/mcb.00396-10.

Yusufzai, T. M., Tagami, H., Nakatani, Y. and Felsenfeld, G. (2004) 'CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species', *Molecular Cell*, 13(2), pp. 291–298. doi: 10.1016/S1097-2765(04)00029-2.

Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. and Flicek, P. R. (2015) 'The Ensembl Regulatory Build', *Genome Biology*, 16(1), pp. 1–8. doi: 10.1186/s13059-015-0621-5.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E. G., Bl??thgen, N., Stadler, M., Tiana, G. and Giorgetti, L. (2017) 'Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes', *Genome Research*, 27(3), pp. 479–490. doi: 10.1101/gr.212803.116.

Zhang, H., Gan, H., Wang, Z., Lee, J. H., Zhou, H., Ordog, T., Wold, M. S., Ljungman, M. and Zhang, Z. (2017) 'RPA Interacts with HIRA and Regulates H3.3 Deposition at Gene Regulatory Elements in Mammalian Cells', *Molecular Cell*, 65(2), pp. 272–284. doi: 10.1016/j.molcel.2016.11.030.

Zhang, T., Paulson, J. R., Bakhrebah, M., Kim, J. H., Nowell, C., Kalitsis, P. and Hudson, D. F. (2016) 'Condensin I and II behaviour in interphase nuclei and cells undergoing premature chromosome condensation', *Chromosome Research*, 24(2), pp. 243–269. doi: 10.1007/s10577-016-9519-7.

Zhang, T., Wei, G., Millard, C. J., Fischer, R., Konietzny, R., Kessler, B. M., Schwabe, J. W. R. and Brockdorff, N. (2018) 'A variant NuRD complex containing PWWP2A/B excludes MBD2/3 to regulate transcription at active genes', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-06235-9.

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., van de Corput, M. P. C., van de Werken, H. J. G., Knoch, T. A., van IJcken, W. F. J., Grosveld, F. G., Ren, B. and Wendt, K. S. (2014) 'Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells', *Proceedings of the National Academy of Sciences*, 111(3), pp. 996–1001. doi: 10.1073/pnas.1317788111.

8 Appendix

8.1 ChIP-Seq targets

Downloaded from GEO:

Target	Author	ID
AFF4	Gardini_A	SRR1342222
AFF4	Gardini_A	SRR1342226
BAP18	Vermeulen_M	SRR060524
BAP18	Vermeulen_M	SRR060530
BDP1	Barski_A	SRR039063
BRD4	Liu_W	SRR1016020
BRF1	Oler_AJ	SRR036646
BRF1	Oler_AJ	SRR036647
BRF2	Oler_AJ	SRR036648
ELL2	Gardini_A	SRR1342230
ELL2	Gardini_A	SRR1342234
EMSY	Varier_RA	SRR2968329
FOXK1	Grant_GD	SRR518708
FOXK1	Grant_GD	SRR518709
GABPA	Michaud_J	SRR332363
GATAD1	Vermeulen_M	SRR060522
GATAD1	Vermeulen_M	SRR060528

GATAD1	Varier_RA	SRR2968331
GATAD2A	Spruijt_CG	SRR3321922
GATAD2B	Spruijt_CG	SRR3321923
GTF2B	Liang_K	SRR2149249
GTF2B	Liang_K	SRR2149275
GTF3C1	Oler_AJ	SRR036662
GTF3C1	Oler_AJ	SRR036663
GTF3C1	Barski_A	SRR039064
H3ac	Liu_W	SRR1016021
H3K27ac	Lai_F	SRR2002376
H3K27ac	Mendoza-Parra_M	SRR3090661
H3K27me3	Vermeulen_M	SRR060537
H3K27me3	Zhong_J	SRR5998761
H3K27me3	Zhong_J	SRR5998762
H3K27me3	Zhong_J	SRR5998763
H3K27me3	Zhong_J	SRR5998764
H3K27me3	Zhong_J	SRR5998765
H3K4me1	Kuznetsova_T	SRR1593537
H3K4me1	Nilson_KA	SRR5195229
H3K4me1	Nilson_KA	SRR5195230
H3K4me2	Liu_W	SRR058521
H3K4me3	Liang_K	SRR2149248

H3K4me3	Liang_K	SRR2149272
H3K79me3	Li_Y	SRR1548387
H3K9ac	Li_Y	SRR1548391
H3K9me3	Tchasovnikarova_IA	SRR1646218
H4ac	Liu_W	SRR1016022
HCFC1	Michaud_J	SRR332357
HJURP	Athwal_RK	SRR1784850
HJURP	Athwal_RK	SRR1784851
LIN9	Sadasivam_S	SRR097890
LRWD1	Vermeulen_M	SRR060534
MBD3	Spruijt_CG	SRR3321918
MBD3	Spruijt_CG	SRR3321919
MYBL2	Sadasivam_S	SRR097892
MYC	Walz_S	SRR765746
P300	Kuznetsova_T	SRR1593543
PHF8	Vermeulen_M	SRR060526
PHF8	Vermeulen_M	SRR060532
POL2RA	Walz_S	SRR765748
POLR2A	Liang_K	SRR2149274
POLR2B	Michaud_J	SRR332367
PWWP2A	Pünzeler_S	SRR3175293
PWWP2A	Pünzeler_S	SRR3175294
RELA	Rao_NA	SRR067999
RPA1	Zhang_H	SRR3091731
RPA2	Zhang_H	SRR3091729

RPA2	Zhang_H	SRR3091730
SGF29	Vermeulen_M	SRR060523
SGF29	Vermeulen_M	SRR060529
TASOR	Tchasovnikarova_IA	SRR5296141
THAP1	Michaud_J	SRR332362
YY1	Zhang_WJ	SRR2059425
YY1	Michaud_J	SRR332364
ZBTB17	Walz_S	SRR765747
ZNF143	Michaud_J	SRR332361

Downloaded from ENCODE:

Target	Author	ID
CTCF	Bernstein_BE	ENCFF474UIC
CTCF	Bernstein_BE	ENCFF163WNN
EZH2	Bernstein_BE	ENCFF162NDI
EZH2	Bernstein_BE	ENCFF081LSC
H2AFZ	Bernstein_BE	ENCFF185PMC
H2AFZ	Bernstein_BE	ENCFF645GVZ
H3K4me3	Bernstein_BE	ENCFF821BHR
H3K4me3	Bernstein_BE	ENCFF093KYN
H3K4me3	Bernstein_BE	ENCFF286DPW
H3K4me3	Bernstein_BE	ENCFF185UFB
H3K27ac	Bernstein_BE	ENCFF440OOW
H3K27ac	Bernstein_BE	ENCFF475KTC
TCF7L2	Farnham_PJ	ENCFF000XMI
TCF7L2	Farnham_PJ	ENCFF000XMJ
POLR2A	Myers_RM	ENCFF452GJT
POLR2A	Myers_RM	ENCFF713XOH
CHD2	Snyder_MP	ENCFF000XBT
CHD2	Snyder_MP	ENCFF000XBW
DEK	Snyder_MP	ENCFF865GUF
DEK	Snyder_MP	ENCFF312DGQ
EP300	Snyder_MP	ENCFF000XJH
EP300	Snyder_MP	ENCFF000XJJ
GTF2F1	Snyder_MP	ENCFF000XEF

GTF2F1	Snyder_MP	ENCFF000XEG
Input2	Snyder_MP	ENCFF586NUV
Input2	Snyder_MP	ENCFF386EWJ
Input2	Snyder_MP	ENCFF456PZV
Input	Snyder_MP	ENCFF316UXC
Input	Snyder_MP	ENCFF939XHK
KAT2A	Snyder_MP	ENCFF000XEB
KAT2A	Snyder_MP	ENCFF000XED
MAFF	Snyder_MP	ENCFF226KER
MAFF	Snyder_MP	ENCFF002SPV
RAD21	Snyder_MP	ENCFF000XKH
RAD21	Snyder_MP	ENCFF000XKI
SMC3	Snyder_MP	ENCFF000XLA
SMC3	Snyder_MP	ENCFF000XLC
TBP	Snyder_MP	ENCFF000XLX
TBP	Snyder_MP	ENCFF000XLW
UBTF	Snyder_MP	ENCFF008CNQ
UBTF	Snyder_MP	ENCFF850IFD
USF2	Snyder_MP	ENCFF000XNC
USF2	Snyder_MP	ENCFF000XND
ZHX1	Snyder_MP	ENCFF917NVM
ZHX1	Snyder_MP	ENCFF604LKK
H3K4me3	Stamatoyannopoulos_J	ENCFF871ZRW
H3K4me4	Stamatoyannopoulos_J	ENCFF959SHX
NFYB	Struhl_K	ENCFF000XIS

NFYB	Struhl_K	ENCFF000XIT
ZZZ3	Struhl_K	ENCFF000XOF
ZZZ3	Struhl_K	ENCFF000XOH

8.2 List of abbreviations

AID	Auxin induced degron
ATAC-Seq	Assay of transposase-accessible chromatin using sequencing
Bait	Baited restriction fragment containing a promoter
bp	base pairs
CHiC	Capture Hi-C
ChIP-Seq	Chromatin immunoprecipitation and sequencing
CRISPR	Clustered regularly interspaced short palindromic repeats
ENCODE	Encyclopaedia of DNA elements
FDR	False discovery rate
GO	Gene ontology
GRCh37	Genome reference consortium human reference genome 37
H*x*K*y*me*z*	Histone H*x* lysine *y* *z*-methylation (e.g. H3K9me1)
H*x*K*y*ac	Histone H*x* lysine *y* acetylation (e.g. H3K9ac)
HAT	Histone acetyl transferase
kb	Kilobases
mRNA	messenger RNA
PCHiC	Promoter capture Hi-C
PIC	Preinitiation complex
PIR	Promoter Interacting Region
Pol II	RNA polymerase 2
RNA pol	RNA polymerase
QC	Quality control
Shh	Sonic hedgehog
TAD	Topologically associating domain
TF	Transcription factor

TSS	Transcription start site

8.3 Supplementary figures

The following figures show a comparison between the approach of calculating ChIP-Seq scores per RF versus the ChIP-Seq peak caller MACS2.

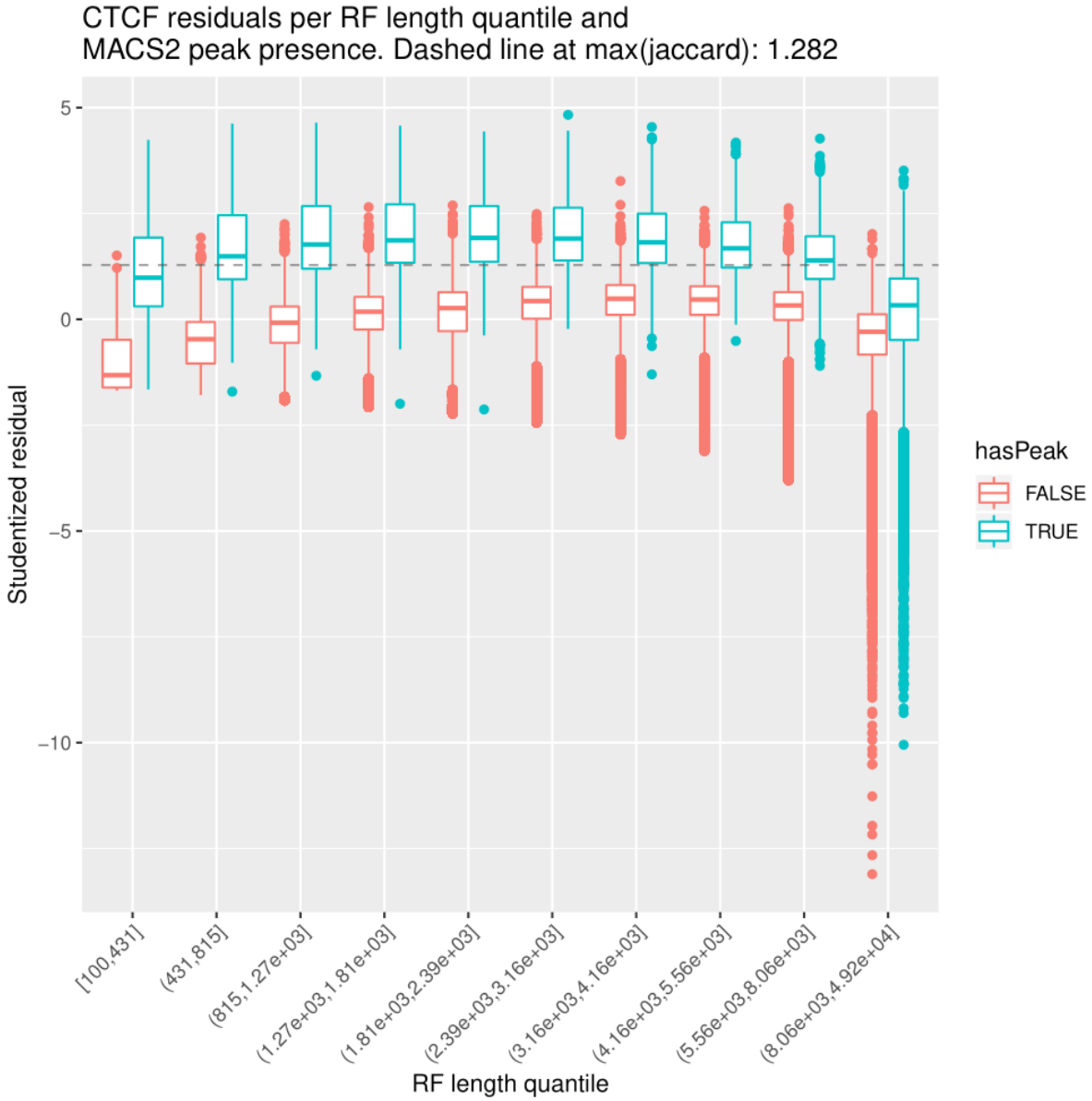

Figure S 1. Comparison between the CTCF ChIP-Seq score per RF and MACS2 peaks. RFs have been stratified by presence of ≥1 MACS2 peak (hasPeak) and by the RF length (x-axis). The y-axis shows the ChIP-Seq score per RF distribution. RFs containing ≥1 MACS2 peak show considerably higher ChIP-Seq scores per RF.

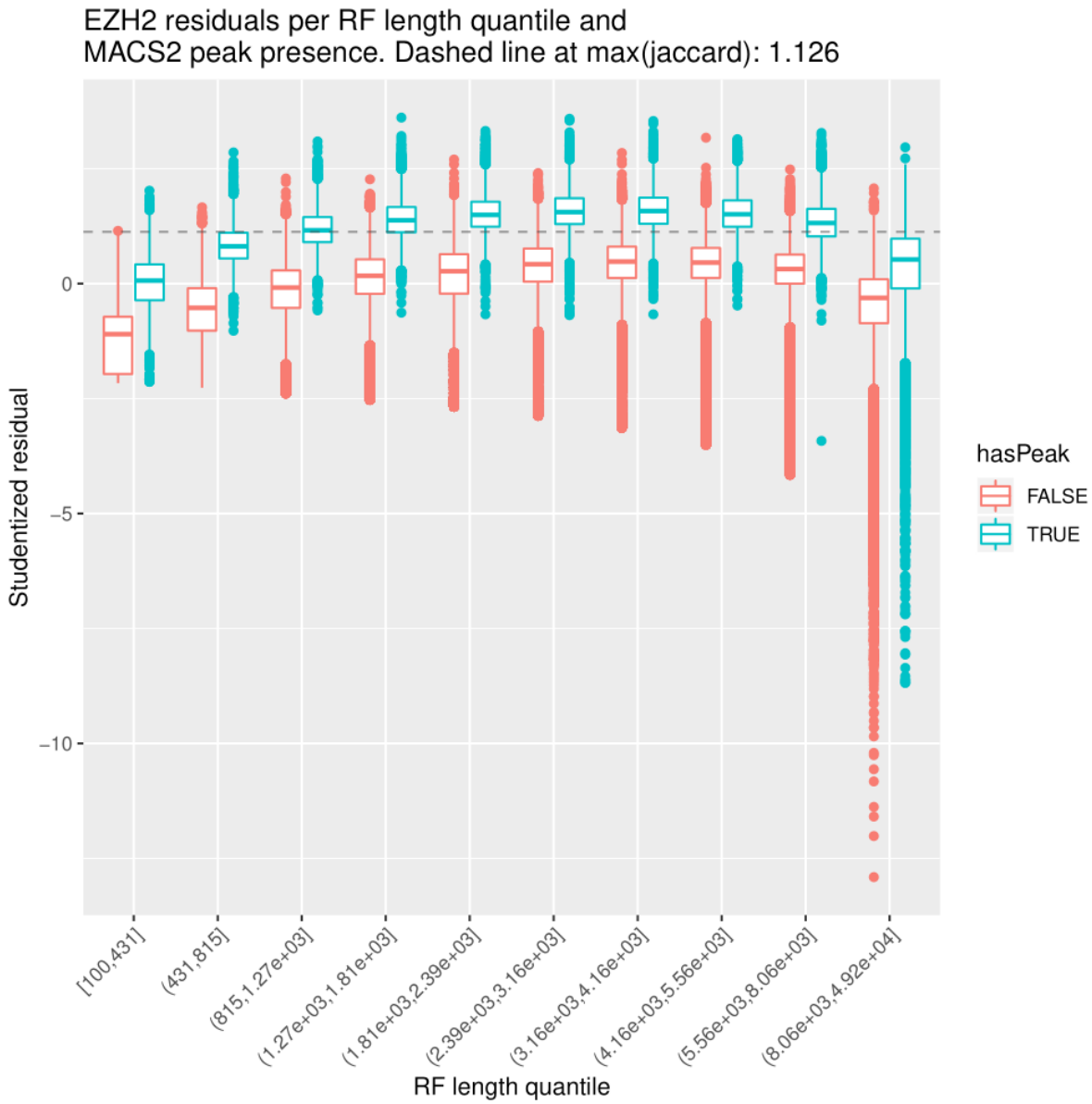

Figure S 2. Comparison between the EZH2 ChIP-Seq score per RF and MACS2 peaks. RFs have been stratified by presence of ≥1 MACS2 peak (hasPeak) and by the RF length (x-axis). The y-axis shows the ChIP-Seq score per RF distribution. RFs containing ≥1 MACS2 peak show considerably higher ChIP-Seq scores per RF.

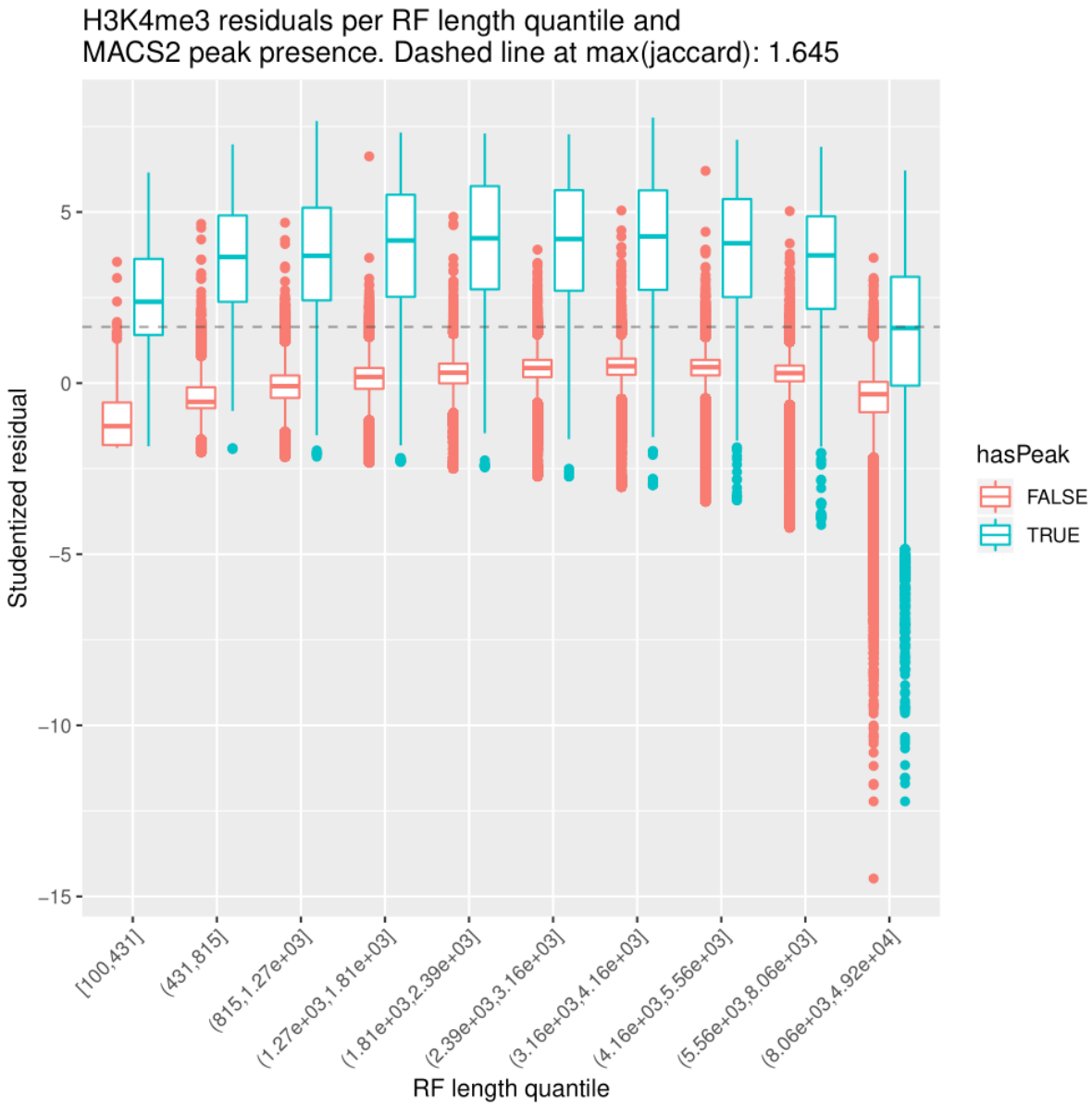

Figure S 3. Comparison between the H3K4me3 ChIP-Seq score per RF and MACS2 peaks. RFs have been stratified by presence of ≥1 MACS2 peak (hasPeak) and by the RF length (x-axis). The y-axis shows the ChIP-Seq score per RF distribution. RFs containing ≥1 MACS2 peak show considerably higher ChIP-Seq scores per RF.

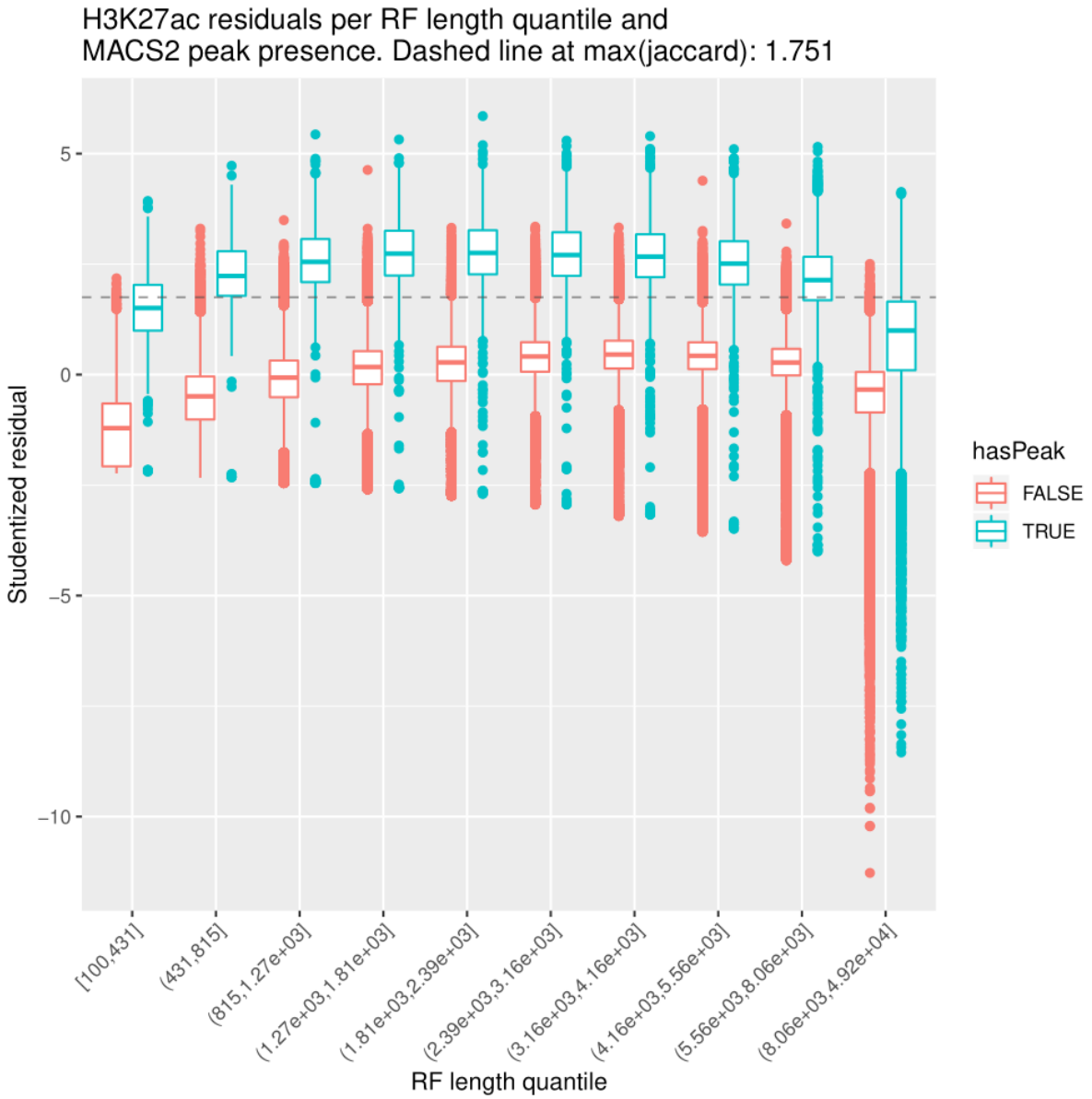

Figure S 4. Comparison between the H3K27ac ChIP-Seq score per RF and MACS2 peaks. RFs have been stratified by presence of ≥1 MACS2 peak (hasPeak) and by the RF length (x-axis). The y-axis shows the ChIP-Seq score per RF distribution. RFs containing ≥1 MACS2 peak show considerably higher ChIP-Seq scores per RF.